

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 29, 2005, 23:18:18 ; Search time 2993.55 Seconds  
(without alignments)  
1990.942 Million cell updates/sec

Title: US-10-657-740-1\_COPY\_51\_173

Perfect score: 639

Sequence: 1 SLFRTVLDGISEVSRDRK.....HAERAIPVSRREKPTSAFSS 123

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2.1/USPTO.spool/US10657740/runat.27052005.165253.3315/app.query.fasta\_1.590  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOFC=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=150  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=50 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10657740 @CGN 1.1.4362 @runat.27052005.165253.3315 -NCPU=6 -ICPU=3  
-NO.MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sv.\*

13: gb.un.\*

14: gb.vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	639	100.0	519	9	CR407691 Homo sapi
2	639	100.0	614	9	BC069528 Homo sapi
3	639	100.0	741	9	U66584 Human alpha
4	639	100.0	1112	9	U05569 Human alpha

5	639	100.0	1114	6	CQ731849
6	602	94.2	543	4	OCRYAA
7	599	93.7	1025	10	MUSCRYAB
8	599	93.7	1056	10	RNU47922
9	598	93.6	793	4	BOVCYA
10	591	92.5	885	10	RNACHY
11	577.5	90.4	796	10	MMU310308
12	577.5	90.4	1271	10	RNU47921
13	550	86.1	447	4	AJ617725
14	548	85.8	447	5	AJ617726
15	544	85.1	522	5	RCACRYST
16	528.5	82.7	558	4	AJ617724
17	519	81.2	765	5	RVCRYA
18	507	79.3	1361	5	D86299
19	505	79.0	447	5	AJ617727
20	502	78.6	448	5	TSU31938
21	492	77.0	705	5	D88185
22	487	76.2	448	5	EUDLAACRY5
23	486	76.1	448	5	APRAAC
24	476	74.5	448	5	CLRAAC
25	474	74.2	730	5	AY035778
26	474	74.2	1430	5	BC083177
27	466	72.9	676	5	AY007972
28	408	63.8	436	5	OLJ000940
29	404	63.2	249	10	RATACRYA
30	401	62.8	249	10	RATACRYB
31	367	57.4	419	9	AF026952
32	367	57.4	148179	9	AP001631
33	367	57.4	163962	9	BS000233
34	367	57.4	340000	9	AP001748
35	364	57.0	507	4	AJ617732
36	364	57.0	632	4	AF029793
37	361	56.5	671	10	S74229
38	361	56.5	687	10	RRHARTABC
39	361	56.5	689	10	RRELENSABC
40	361	56.5	704	10	S77142
41	361	56.5	706	10	S77138
42	361	56.5	1247	6	AX401743
43	361	56.5	1247	10	RATCRYAB
44	358	56.0	548	4	OCRYAB
45	357	55.9	511	10	SEH272441
46	357	55.9	528	10	SJU293658
47	356	55.7	537	5	RCACRYST
48	356	55.7	522	10	HAMSCRYEB
49	354	55.4	528	9	BT0006770
50	354	55.4	528	12	BT007909
51	354	55.4	537	6	AX937702
52	354	55.4	537	6	AX937703
53	354	55.4	691	6	CQ812349
54	354	55.4	691	6	AX330255
55	354	55.4	691	6	AX330332
56	354	55.4	691	6	BD079402
57	354	55.4	691	9	S45630
58	354	55.4	744	9	RC007008
59	354	55.4	856	6	AX888028
60	354	55.4	856	6	BD027638
61	354	55.4	893	6	AX899079
62	354	55.4	893	6	BD034612
63	354	55.4	911	6	AX899075
64	354	55.4	911	6	BD034608
65	354	55.4	913	6	AX899074
66	354	55.4	913	6	BD034607
67	354	55.4	927	6	AX899073
68	354	55.4	927	6	BD034606
69	354	55.4	942	6	AX888027
70	354	55.4	942	6	BD027637
71	354	55.4	1036	6	BD135185
72	354	55.4	1036	6	AX017495
73	352	55.2	733	9	AB125159
74	352	55.1	666	10	MUSCRYABA
75	352	55.1	848	10	BC010768
76	352	55.1	966	5	BC082122
77	350	54.8	661	9	AF007162

c

c

c

78	343	53.7	689	5	DUKABC	L08078 Anas platyr	RESULT 1	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
79	343	53.7	906	10	MAACRY2	X02951 Hamster alp	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
80	342	53.5	846	5	GGU26661	U26661 Gallus gall	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
81	342	53.5	1042	5	BX930014	BX930014 Gallus ga	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
82	342	53.5	1251	5	S53164	S53164 alpha B-cry	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
83	340	53.2	167475	10	AC090881	AC090881 Mus muscu	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
84	340	53.2	188810	2	AC121138	AC121138 Mus muscu	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
85	340	53.2	222895	2	AC084065	AC084065 Mus muscu	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
86	339	53.1	466	4	AJ617731	AJ617731 Macropus	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
87	339	53.1	900	5	AF159089	AF159089 Danio rer	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
88	332	52.0	466	4	AJ617730	AJ617730 Digeiphis	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
89	330	52.0	502	4	AJ617729	AJ617729 Tachyglos	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
90	332	51.6	577	5	AY007973	AY007973 Clarias b	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
91	330	51.6	695	6	AX888029	AX888029 Sequence	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
92	330	51.6	695	6	AX888029	AX888029 Sequence	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
93	329	51.5	501	4	AJ617728	AJ617728 Ornithorh	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
94	321	50.2	1316	10	SEHCYAA2	MI7249 Mole rat al	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
95	320	50.1	1955	5	BC076518	BC076518 Danio rer	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
96	317.5	49.7	765	5	RTCRVA	X00716 Frog mRNA f	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
97	310.5	48.6	888	5	CHKRYAA	BC075197 Xenopus l	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
98	310	48.5	6311	5	CHKRYAA	MI7627 Chicken alp	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
99	309	48.4	826	6	AR415535	AR415535 Sequence	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
100	309	48.4	826	6	AR415535	AR415535 Sequence	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
101	309	48.4	826	6	AR415535	AR415535 Sequence	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
102	293	45.9	168	6	HUMAAAC	BD111088 EST and e	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
103	282.5	44.2	373	6	CO681472	L25781 Homo sapien	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
104	275.5	43.1	1065	5	AY184812	CO681472 Sequence	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
105	273.5	42.8	383	11	BV209856	AY184812 Clarias b	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
106	271	42.4	830	3	AF237691	BV209856 CRVA 464	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
107	268	41.9	787	10	CLSHSP	AF237691 Clona Int	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
108	267	41.8	763	5	GGIAP	X51747 Cricetulus	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
109	266.5	41.7	165693	5	BX248514	X59541 Chicken mRN	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
110	266.5	41.7	166831	2	BX950201	BX248514 Zebrafish	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
111	263	41.2	349	6	AX786923	AX950201 Danio ter	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
112	263	41.2	882	10	MUSHSP25PS	AX786923 Sequence	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
113	262	41.0	1085	5	BC078509	Li1610 Mus musculu	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
114	261.5	40.9	615	9	CR407614	BC078509 Xenopus l	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
115	261.5	40.9	618	9	CR536489	CR407614 Homo sapi	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
116	261.5	40.9	640	9	HSW800252	CR536489 Homo sapi	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
117	261.5	40.9	724	9	BC014920	AL050380 Homo sapi	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
118	261.5	40.9	764	6	CQ799993	BC014920 Homo sapi	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
119	261.5	40.9	764	9	AB020027	CQ799993 Sequence	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
120	261.5	40.9	781	6	AB020027	AB020027 Homo sapi	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
121	261.5	40.9	789	6	AR380755	BC012768 Homo sapi	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
122	261.5	40.9	789	9	HSRSP27L	AR380755 Sequence	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
123	261.5	40.9	794	9	BC073768	X54079 Human mRNA	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
124	261.5	40.9	847	6	BD186273	BC073768 Homo sapi	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
125	261.5	40.9	865	6	CQ730135	BD186273 STAT6 act	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
126	261.5	40.9	865	9	HSU90906	CQ730135 Sequence	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
127	261.5	40.9	867	9	BC000510	U90906 Human clone	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
128	261.5	40.9	1231	6	AX411221	BC000510 Homo sapi	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
129	261.5	40.9	1380	6	AR217508	AX411221 Sequence	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
130	261.5	40.9	1380	6	AR217508	Z23090 H. sapiens m	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
131	261.5	40.9	1380	6	AR217508	AR217508 Sequence	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
132	261	40.8	604	6	AX401305	AX401305 Sequence	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
133	260	40.7	535	10	MMU03562	U03562 Mus musculu	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
134	260	40.7	601	6	MMU03561	U03561 Mus musculu	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
135	260	40.7	618	6	A76413	A76413 Sequence 1	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
136	260	40.7	634	10	MMU03560	U03560 Mus musculu	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
137	260	40.7	787	6	CQ858777	CQ858777 Sequence	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
138	260	40.7	787	6	AX401752	AX401752 Sequence	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
139	260	40.7	787	10	RATHSP27A	M86389 Rattus norv	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
140	260	40.7	862	10	BC018257	BC018257 Mus muscu	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
141	260	40.7	1379	6	AR175769	AR175769 Sequence	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
142	260	40.7	1379	6	AR232004	AR232004 Sequence	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
143	260	40.7	1379	6	AX136008	AX136008 Sequence	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
144	260	40.7	1379	6	AX136777	AX136777 Sequence	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
145	260	40.7	1379	6	AX137828	AX137828 Sequence	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
146	260	40.7	1379	6	BD000152	BD000152 Process f	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
147	260	40.7	1379	6	BD010884	BD010884 Process f	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
148	259.5	40.6	832	5	BC078511	BC078511 Xenopus l	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
149	259.5	40.6	4321	5	AFACCEP1G	BC078511 Xenopus l	LOCUS	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004
150	259	40.5	228912	2	AC114562	AC114562 Mus muscu	DEFINITION	CR407691	519 bp	mrna	linear	PRI 10-MAY-2004

## ALIGNMENTS

RESULT 1  
CR407691  
LOCUS

DEFINITION  
CR407691  
519 bp  
mrna  
linear  
PRI 10-MAY-2004  
Homo sapiens full open reading frame cDNA clone RZPD0834E043D for  
gene CRVAA, crystallin, alpha A complete cds, without stopcodon.

ACCESSION  
CR407691  
VERSION  
CR407691.1  
KEYWORDS  
Full ORF shuttle clone, Gateway(TM), complete cds.  
SOURCE  
Homo sapiens (human)

ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
1 (bases 1 to 519)  
AUTHORS  
Ebert,L., Schick,M., Neubert,P., Schatten,R., Henze,S. and Korn,B.

TITLE  
Cloning of human full open reading frames in Gateway(TM) system  
entry vector (pDONR201)

JOURNAL  
Unpublished

REFERENCE  
2 (bases 1 to 519)  
AUTHORS  
Ebert,L., Schick,M., Neubert,P., Schatten,R., Henze,S. and Korn,B.

TITLE  
Direct Submission

JOURNAL  
Submitted (07-MAY-2004) RZPD Deutsches Ressourcenzentrum fuer  
Genomforschung GmbH Im Neuenheimer Feld 580, D-69120 Heidelberg,  
Germany

COMMENT  
RZPD: RZPD0834E043D, ORFNO 642  
www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=RZPD0834E043D RZPDLIB;  
Human Full ORF Clones Gateway(TM) - RZPD (kan-resist.) RZPD LIB No.  
834

www.rzpd.de/cgi-bin/products/showLib.pl.cgi/response?libNo=834  
www.rzpd.de/products/orfclones/  
Contact: Ina Rolfs

RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
Heubnerweg 6, D-14059 Berlin, Germany  
Tel: +49 30 32639 101  
Fax: +49 30 32639 111  
www.rzpd.de

This clone is available from RZPD;  
contact RZPD (customer.service@rzpd.de) for further information.

This CDS clone is a part of a collection of human full length  
expression clones generated by RZPD.

This CDS has been inserted into pDONR201 via a BP Clonase(TM)  
reaction. Additional sequence has been added in front of the start  
codon (ATG): att. AAAAAA GCT GGC ACC CCT GGT CCA GGT (ATG)

After the last codon additional sequence has been added: CCA GGC  
CCA GGC GGC G in front of the 3' att site (AC CCA GCT TTC TT).

Compared to the reference sequence U66584 we did not find any amino  
acid exchanges.

Clone distribution: http://www.rzpd.de/products/orfclones/.  
Location/Qualifiers

1. 519  
/organism="Homo sapiens"  
/mol\_type="mrna"

/db\_xref="taxon:9606"  
/clone="RZPD0834E043D"

/clone\_lib="Human Full ORF Clones Gateway(TM) - RZPD"  
/lab\_host="DH10B"

/note="Vector: pDONR201, Site\_1: attP1; Site\_2: attP2"  
1. 519  
/gene="CRYAA"

1. >519  
/gene="CRYAA"

/codon\_start=1  
/protein\_id="CAG28619.1"

/db\_xref="GI:47115319"

/translation="MDVTIQHWFKRTLGGPYPSRLPQDFGEGFLFEVDLLPFLSSTI  
SPYRQSLFRVLDSGISEVSRDVKFVFLDVKHFSPELTVKQDDFVEIHGKNE  
RQDDHGYISREHFRYLRPLSNVDQSALSCSLADGMLTFCGPKIQTLGDLTAHERAIP  
VSREKPTSPSS"

gene

CDS

ORIGIN

## Alignment Scores:

Pred. No.: 3, 77e-63 Length: 519  
 Score: 639.00 Matches: 123  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 9 Gaps:

US-10-657-740-1\_COPY\_51\_173 (1-123) x CR407691 (1-519)

Qy 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20  
 Db 151 TCCCTCTTCCGACCGTCTGACTCCGGCATCTCTGAGTTCGATCCGACCGGACAG 210  
 Qy 21 PheValIlePheLeuAspValIleHisPheSerProGluAspLeuThrValIleValGln 40  
 Db 211 TTCGTGATCTCTCTGATGTGAAGCACCTTCTCCCGGAGGACCTCACCGTGAAGGTGAC 270  
 Qy 41 AspAspPheValGluIleHisGlyIleHisGlnGluArgGlnAspAspHisGlyTyrIle 60  
 Db 271 GACGACTTTGTGGAGATCCAGGAAACACACGAGCGCCAGGACGACCGGTACATT 330  
 Qy 61 SerArgGluPheHisArgArgTyrArgLeuProSerAnValAspGlnSerAlaLeuSer 80  
 Db 331 TCCCGTAGTTCCACCGCGCTACCGCTCGCTCCAAACGGAGACCGTCCGCTCTCT 390  
 Qy 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100  
 Db 391 TGCTCCCTGTTCTGCGGATGGCATGCTGACCTTCTGTGCCCCAAGATCCAGACTGGCCTG 450  
 Qy 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120  
 Db 451 GATGCCACCCACGCGGAGCGCATCCCGTGTGCGGGAGGAGAGCCACCTCGGCT 510  
 Qy 121 ProSerSer 123  
 Db 511 CCCTCGTCC 519

## RESULT 2

BC069528 Homo sapiens crystallin, alpha A, mRNA (cdna clone MGC:96924  
 IMAGE:7262133), complete cds.  
 BC069528  
 VERSION  
 KEYWORDS MGC.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

## REFERENCE

1 (Bases 1 to 614)  
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheerz, T.E., Brownstein, M.J., Usdin, T.B., Toehiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullah, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smalilus, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

## TITLE

## JOURNAL

12477932

2 (bases 1 to 614)

REFERENCE Strausberg, R.

AUTHORS Direct Submission

TITLE Submitted (29-APR-2004) National Institutes of Health, Mammalian

JOURNAL Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov

COMMENT Contact: MGC help desk

Email: cgapbs-remail.nih.gov

Tissue Procurement: Baylor Human Genome Sequencing Center

cDNA Library Preparation: Baylor Human Genome Sequencing Center

cDNA Library Arrayed by: The I.M.A.G.E.-E. Consortium (LLNL)

DNA Sequencing By: Baylor College of Medicine Human Genome

Sequencing Center

Center code: BCM-HGSC

Web site: http://www.hgsc.bcm.tmc.edu/cdna/

Contact: amg@bcm.tmc.edu

Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,

Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,

A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E.-E. Consortium/LLNL at: http://image.llnl.gov  
 Series: IRBR Plate: 1 Row: d Column: 7.

## FEATURES

## source

1..614  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="MGC:96924 IMAGE:7262133"  
 /tissue type="PCR rescued clones"  
 /clone lib="NIH MGC 244"  
 /note="vector: pPCR-Script Amp SK(+)"

## gene

1..614  
 /gene="CRYAA"  
 /note="synonyms: CRYA1, HSPB4"  
 /db\_xref="LocusID:1409"  
 /db\_xref="MIM:123580"  
 51..572  
 /gene="CRYAA"

## CDS

/codon\_start=1  
 /product="crystallin, alpha A"  
 /protein\_id="AAH69528.1"  
 /db\_xref="GI:46854599"  
 /db\_xref="LocusID:1409"  
 /db\_xref="MIM:123580"  
 /translation="MDVTIQHPFKRTLGPYPSRLFDQFGEGLFEYDLLPFLSSTI  
 SPYRQSLFTVLDSGISSEVRSDRDKFVFLDVKHFPEDLTVKVQDDFVEIHGKNE  
 RQDDHGVISREFFHRRYRLPSNVDSALSCSLSDGMLTFCGPKIQTLGLDTHAERAIP  
 VSREKPTSPSS"

## ORIGIN

Alignment Scores:  
 Pred. No.: 4, 59e-63 Length: 614  
 Score: 639.00 Matches: 123  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 9 Gaps:

US-10-657-740-1\_COPY\_51\_173 (1-123) x BC069528 (1-614)

Qy 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20  
 Db 201 TCCCTCTTCCGACCGTCTGACTCCGGCATCTCTGAGTTCGATCCGACCGGACAG 260  
 Qy 21 PheValIlePheLeuAspValIleHisPheSerProGluAspLeuThrValIleValGln 40  
 Db 261 TTCGTGATCTCTCTGATGTGAAGCACCTTCTCCCGGAGGACCTCACCGTGAAGGTGAC 320  
 Qy 41 AspAspPheValGluIleHisGlyIleHisGlnGluArgGlnAspAspHisGlyTyrIle 60

```

|||||
321 GACGACTTTGTGGAGATCCAGGAAGCACAAAGAGCCAGGACGACCAACGGCTACATT 380
QY 61 SerArgGluPheHisArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80
Db 381 TCCCGTGAAGTTCCACCGCGCTACCGCTCGCGTCCAAACGTGGACCACTCGGCCCTCTCT 440
QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100
Db 441 TCTCTCTCTGTGGCGATGGCATGCTGACCTTCTGTGCCCCCAAGATCCAACTGGCCCTG 500
QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120
Db 501 GATGCCACCCACCGCGAGGACCATCCCGTGTGCGGGAGGAGAGACCCACCTCGGCT 560
QY 121 ProSerSer 123
Db 561 CCCTCGTCC 569

RESULT 3
HSU66584
LOCUS Human alphaA-crystallin (CRYAA) mRNA, complete cds. PRI 14-DEC-1996
DEFINITION
ACCESSION U66584
VERSION U66584.1 GI:1732062
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS Petrash,J.M., Mathur,S., Manoharan,M. and Andley,U.P.
TITLE Cloning and expression of human lens crystallins
JOURNAL Invest. Ophthalmol. Vis. Sci. 36, S882-S882 (1995)
REFERENCE
AUTHORS Andley,U.P., Mathur,S., Griest,T.A. and Petrash,J.M.
TITLE Cloning, expression, and chaperone-like activity of human
alphaA-crystallin
JOURNAL J. Biol. Chem. 271 (50), 31973-31980 (1996)
MEDLINE 97112991
PUBMED 8943244
REFERENCE
AUTHORS Petrash,J.M., Mathur,S., Wang,J.C., Griest,T.A. and Andley,U.P.
TITLE Direct Submission
JOURNAL Submitted (12-AUG-1996) Ophthalmology and Visual Sciences,
Washington University School of Medicine, 660 S. Euclid Ave., St.
Louis, MO 63110, USA

FEATURES
source
location/Qualifiers
1..741
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="eye lens"
1..741
/gene="CRYAA"
1..522
/function="structural protein"
/note="structurally related to sHSPs; chaperone-like
activity"
/codon_start=1
/product="alphaA-crystallin"
/protein_id="AAC50900.1"
/db_xref="GI:1732063"
/translation="MDVTIQHFWFKRTIGPFYPSRLFDQFFGEGLEFYEDLLPFLSSTI
SPYKQSLPFTVLDGISEVSRDRKPFVFLDVKHPSFEDLVKQDDFVEIHGKNE
RQDDHGYSIREPHRYRLPSNDVQSALSCSLSDGMLTFCGPKIQTLGDATHAERAIP
VSREKPTSPSS"

ORIGIN
Alignment Scores: 5.72e-63 Length: 741
Pred. No.: 639.00 Matches: 123
Score:

```

```

Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Db: 9 Gaps: 0
US-10-657-740-1_COPY_51_173 (1-123) x HSU66584 (1-741)

QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20
Db 151 TCCCTCTTCGACCGGTGGACTCGGCATCTCTGAGGTTGATCCGACCGGACAAG 210
QY 21 PheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrValLysValGln 40
Db 211 TTCGTATCTTCTCGATGTGAAGCACTTCTCCCGGAGGACCTCACCGTGAGGTGCAG 270
QY 41 AspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIle 60
Db 271 GACGACTTTGTGGAGATCCACGGAAGACAAACAGCGCCAGGACCAACGGTACATT 330
QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80
Db 331 TCCGTTGAGTTCCACCGCGCTACCGCTCCAAACGTGGACCACTCGGCCCTCTCT 390
QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100
Db 391 TGCTCCCTGTCTGCCGATGGCATGCTGACCTTCTGTGCCCCCAAGATCCAACTGGCCTG 450
QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120
Db 451 GATGCCACCCACCGCGAGGACCATCCCGTGTGCGGGAGGAGAGACCCACCTCGGCT 510
QY 121 ProSerSer 123
Db 511 CCCTCGTCC 519

RESULT 4
HSU05569
LOCUS Human alphaA-crystallin (CRYA1) mRNA, complete cds. PRI 25-APR-1996
DEFINITION
ACCESSION U05569
VERSION U05569.1 GI:452477
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS Jaworski,C.J.
TITLE The human alphaA-crystallin gene
JOURNAL Thesis (1992) LMDB, NEI, Molecular Structure and Function
REFERENCE 2 (bases 1 to 1112)
AUTHORS Jaworski,C.J.
TITLE A reassessment of mammalian alpha A-crystallin sequences using DNA
sequencing: implications for anthropoid affinities of tarsier
J. Mol. Evol. 41 (6), 901-908 (1995)
MEDLINE 96139023
PUBMED 8587135
REFERENCE 3 (bases 1 to 1112)
AUTHORS Wistow,G.J.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-1994) Graeme J. Wistow, Molecular Structure and
Function, LMDB, NEI, NIH, Bethesda, MD 20892, USA

FEATURES
source
location/Qualifiers
1..1112
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="21"
/tissue_type="lens"
1..1112
/gene="CRYA1"
68..589
/gene="CRYA1"

gene
CDS

```



/function="lens structural protein"  
 /note="This sequence is derived from genomic and cDNA  
 sequence"  
 /citation={1}  
 /citation={2}  
 /codon\_start=1  
 /evidence=experimental  
 /product="alphaA-crystallin"  
 /protein\_id="AAA97523.1"  
 /db\_xref="GI:452478"  
 /translation="MDVTIOHPFKRTLGPPYPSRLFDQFEGEGLEFYDLPLFLSSITI  
 SPYKQSLPRTVDSGISVSDRDKFVILFDVKHSPEDLTIVQDDFVHKGHNE  
 RQDDGYLRSRFBRYRLPSNVDSALSLSADGMLTFCGPKIQTLGLDATAERAIP  
 VSREKPTSAFSS"

## ORIGIN

Alignment Scores:  
 Pred. No.: 9,19e-63 Length: 1112  
 Score: 639.00 Matches: 123  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 9 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x HSU05569 (1-1112)

Qy 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20  
 Db 218 TCCCTCTTCCGACCGTCTGACCTCCGGCATCTCTGAGGTTCCGATCCGCGGACAA 277  
 Qy 21 PheValIlePheLeuAspValIleHisPheSerProGluAspLeuThrValIleValGln 40  
 Db 278 TTGCTCATCTTCTCCGATGTGAAGCACCTTCTCCCGGAGGACCTCACCGTGAAGGTGCAG 337  
 Qy 41 AspAspPheValGluIleHisGlyIleHisGlyIleHisGlyIleHisGlyIleHisGlyIle 60  
 Db 338 GACGACTTTGTGGAGATCCAGAAAGCACACGAGCCGACGAGACGACGACGCTACATT 397  
 Qy 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
 Db 398 TCCCGTGAAGTTCACCGCGCTACCGCTCGCGTCCAGCTGACGACGACGACGACGCTCTCT 457  
 Qy 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100  
 Db 458 TGCTCCCTGTCTCGCGATGGATGTCGACCTTCTGTGCGCCCAAGATCCAGACTGGCCTG 517  
 Qy 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120  
 Db 518 GATGCCACCCAGCGGAGGACCATCCCGTGTCCGCGGAGGAGAGGCCACCTCGGCT 577  
 Qy 121 ProSerSer 123  
 Db 578 CCCTCGTCC 586

RESULT 5  
 LOCUS COQ731849 1114 bp DNA linear PAT 03-FEB-2004  
 DEFINITION Sequence 17783 from Patent WO2068579.  
 ACCESSION COQ731849  
 VERSION COQ731849.1 GI:42309516  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE  
 AUTHORS Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.  
 TITLE Kits, such as nucleic acid arrays, comprising a majority of  
 humanexons or transcripts, for detecting expression and other uses  
 thereof  
 JOURNAL Patent: WO 02068579-A 17783 06-SEP-2002;  
 PE Corporation (NY) (US)  
 FEATURES Location/Qualifiers

## source

1. .1114  
 /organism="Homo sapiens"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"

## ORIGIN

Alignment Scores:  
 Pred. No.: 9,21e-63 Length: 1114  
 Score: 639.00 Matches: 123  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x CQ731849 (1-1114)

Qy 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20  
 Db 220 TCCCTCTTCCGACCGTCTGACCTCCGGCATCTCTGAGGTTCCGATCCGACCGGACAA 279  
 Qy 21 PheValIlePheLeuAspValIleHisPheSerProGluAspLeuThrValIleValGln 40  
 Db 280 TTGCTCATCTTCTCCGATGTGAAGCACCTTCTCCCGGAGGACCTCACCGTGAAGGTGCAG 339  
 Qy 41 AspAspPheValGluIleHisGlyIleHisGlyIleHisGlyIleHisGlyIleHisGlyIle 60  
 Db 340 GACGACTTTGTGGAGATCCAGAAAGCACACGAGCCGACGAGACGACGCTACATT 399  
 Qy 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
 Db 400 TCCCGTGAAGTTCACCGCGCTACCGCTCGCGTCCAGCTGACGACGACGACGACGCTCTCT 459  
 Qy 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100  
 Db 460 TGCTCCCTGTCTCGCGATGGATGTCGACCTTCTGTGCGCCCAAGATCCAGACTGGCCTG 519  
 Qy 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120  
 Db 520 GATGCCACCCAGCGGAGGACCATCCCGTGTCCGCGGAGGAGAGGCCACCTCGGCT 579  
 Qy 121 ProSerSer 123  
 Db 580 CCCTCGTCC 588

RESULT 6  
 LOCUS OCCRYAA 543 bp mRNA linear MAM 23-OCT-1996  
 DEFINITION O.cuniculus mRNA for alpha-A-crystallin.  
 ACCESSION X95382  
 VERSION X95382.1 GI:1177576  
 KEYWORDS alpha-A-crystallin; cryaA gene.  
 SOURCE Oryctolagus cuniculus (rabbit)  
 ORGANISM Oryctolagus cuniculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.

REFERENCE  
 AUTHORS Krausz,E., Augusteyn,R.C., Quinlan,R.A., Reddan,J.R., Russell,P.,  
 Sax,C.M. and Graw,J.  
 TITLE Expression of Crystallins, Pax6, Filensin, CP49, MIP, and MP20 in  
 lens-derived cell lines  
 JOURNAL Invest. Ophthalmol. Vis. Sci. 37 (10), 2120-2128 (1996)  
 MEDLINE 96409169  
 PUBMED 8814151  
 REFERENCE 2 (bases 1 to 543)  
 AUTHORS Graw,J.  
 TITLE Direct Submission  
 JOURNAL Submitted (29-JAN-1996) J. Graw, Institute of Mammalian Genetics,  
 GSF-Research Center Neuherberg, Ingolstaedter Landstrasse 1,  
 Oberschleissheim, D-85764, FRG  
 FEATURES Location/Qualifiers  
 source  
 1. .543  
 /organism="Oryctolagus cuniculus"  
 /mol\_type="mRNA"

gene /sub\_species="New Zealand white"  
/db\_xref="taxon:9986"  
/tissue\_type="lens"

CDS

1..543 /gene="cryaa"  
22..543 /gene="cryaa"  
/codon\_start=1  
/product="alpha-A-crystallin"  
/protein\_id="CAA64668.1"  
/db\_xref="GI:1177577"  
/db\_xref="GOA:P02493"  
/db\_xref="UniProt/Swiss-Prot:P02493"  
/translation="MDVTIOHPFKRTIGPPYPSRLFDQFFGEGFLFYDILLPLFLSSTI  
SPYRQSLPTVLDGSISEVRSDRKDFVFLDVKHFSPEDLTVKVEDFVEIHGKHNE  
RODDHGYISREPHRRYRLPSNVDSALSCSLSDGMLTFSGPKVQSGLDAGHSRAIP  
VSREKPSVPSS"

# ORIGIN

Alignment Scores:  
Pred. No.: 6,27e-59 Length: 543  
Score: 602.00 Matches: 115  
Percent Similarity: 97.56% Conservative: 5  
Best Local Similarity: 93.50% Mismatches: 3  
Query Match: 94.21% Indels: 0  
DB: 4 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x OCCRYAA (1-543)

QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20  
Db 172 TCCCTCTTCCGACCGTGTGACTCCGGCATCTCTGAGTCCGCTCCGACGGGACAAG 231  
QY 21 PheValIlePheLeuAspValIleHisPheSerProGluAspLeuThrValIleValGln 40  
Db 232 TTCTGTCATCTTCTGAGCGTGAAGACATCTCTCCGCGAGGACCTCACCGTGAAGTGCAG 291  
QY 41 AspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIle 60  
Db 292 GAGGACTCTGTGAGATCCACGGCAACACACAGAGACAGATGACCGGTACATT 351  
QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
Db 352 TCCCGCGAGTTCACCGCGCTTACCGCTGCCCTTCCACCTGCACGACGAGTCCGCGCTGCC 411  
QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100  
Db 412 TCTCGCTCTCCGCGACGGCATCTCACCTTCTCCGGCCCCCAAGTGAGTCCGCGCTG 471  
QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120  
Db 472 GACGCTGGCCACAGCGAGCGACCCATCCCGTGTCCCGGAGGAGAGAGCCAGTCCGCTG 531  
QY 121 ProSerSer 123  
Db 532 CCCTCGTCC 540

# RESULT 7

MUSCRYAB 1025 bp mRNA linear ROD 27-APR-1993  
LOCUS Mouse lens alpha-A-crystallin mRNA, 3' end.  
DEFINITION J00376  
ACCESSION J00376.1 GI:192760  
VERSION alpha-A-crystallin; crystallin.  
KEYWORDS alpha-A-crystallin (house mouse)  
SOURCE Mus musculus  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 1025)  
REFERENCE King, C.R., Shinohara, T. and Piatigorsky, J.  
AUTHORS alpha-A-crystallin messenger RNA of the mouse lens: more noncoding  
TITLE than coding sequences  
JOURNAL Science 215 (4535), 985-987 (1982)

MEDLINE 83119896  
PUBMED 7156978  
COMMENT

The crystallins contribute 90% of the soluble lens protein and are highly conserved during evolution. There are four immunologically separate classes of crystallins called alpha-, beta-, gamma-, and delta-crystallin. The mRNA of the alpha-A-crystallin from mouse, like that from cows and rats, is nearly three times the size required to code for the polypeptide. Nucleotides 1-490 of the mouse alpha-A-crystallin mRNA predict a protein sequence identical to amino acids 10-173 of rat alpha-A-crystallin.

# FEATURES

Location/Qualifiers  
source 1..1025  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
mRNA <1..1025  
/product="CRY-alpha-A mRNA"  
CDS <1..492  
/note="alpha-A-crystallin"  
/codon\_start=1  
/protein\_id="AAA37471.1"  
/db\_xref="GI:387134"  
/translation="KRALGPPYPSRLFDQFFGEGFLFYDILLPLFLSSTIPYRQSLFR  
TVLDGSISEVRSDRKDFVFLDVKHFSPEDLTVKVEDFVEIHGKHNERODDHGYISR  
EFHRRYRLPSNVDSALSCSLSDGMLTFSGPKVQSGLDAGHSRAIPVSREKPSA  
PSS"

ORIGIN 70 bp upstream of Real site.

Alignment Scores:  
Pred. No.: 2,89e-58 Length: 1025  
Score: 599.00 Matches: 115  
Percent Similarity: 97.56% Conservative: 5  
Best Local Similarity: 93.50% Mismatches: 3  
Query Match: 93.74% Indels: 0  
DB: 10 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x MUSCRYAB (1-1025)

QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20  
Db 121 TCCCTCTTCCGACCTGTGACTCCGGCATCTCTGAGTCCGCTCCGACGGGACAAG 180  
QY 21 PheValIlePheLeuAspValIleHisPheSerProGluAspLeuThrValIleValGln 40  
Db 181 TTCTGTCATCTTCTGAGCGTGAAGACATCTCTCTCTGAGGACCTCACCGTGAAGTACTG 240  
QY 41 AspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIle 60  
Db 241 GAAGATTGTGTGAGATTACGGCAACACACAGAGAGGACGATGACCTGGCTTACATA 300  
QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
Db 301 TCCGCTGAATTTCACCGTCCGTACCGTCTGCTTCCATGTGGACCACTCCGCCCTCTCC 360  
QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100  
Db 361 TGCTCCCTGTCTGCTGATGCGATGCTGACCTTCTTGCCCCCAAGTCCAGTCCGCTTG 420  
QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120  
Db 421 GATGCTGCCACACAGGAGGCGCATCTCTGTGTCACTGTGTGAGGAGGAGAACCCAGCTCTGCA 480  
QY 121 ProSerSer 123  
Db 481 CCCTCGTCC 489  
RESULT 8  
LOCUS RNU47922  
DEFINITION Rattus norvegicus alpha A-crystallin mRNA, complete cds.  
ACCESSION U47922  
VERSION U47922.1 GI:1245161

**KEYWORDS**  
**SOURCE** Rattus norvegicus (Norway rat)  
**ORGANISM** Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
**REFERENCE** 1 (bases 1 to 1056)  
**AUTHORS** Bhat, S.P., Nandy, P., Srinivasan, A., Cheng, D. and Sitay, A.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (31-JAN-1996) Suraj P. Bhat, JSEI, UCLA School Medicine,  
 100-Stein Plaza, Rm. B623, Los Angeles, CA 90095-7008, USA  
**FEATURES**  
 Location/Qualifiers  
 1..1056  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /strain="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 13..534  
 /note="similar to small heat shock proteins"  
 /codon\_start=1  
 /product="alpha A-crystallin"  
 /protein\_id="AAA93367.1"  
 /db\_xref="GI:1245162"  
 /translation="MDVIGIQHPWFKRALGPVPSRLFOQFGEGLFEYDLPLFLSSTI  
 SPYRQSLFRVLDSGISSEVSRDRDKFVILDKHFSPEDLTVKVLDFVEIHGKNE  
 RQDDHGYISREHFHRRYRLPSNVDSQALSCLSDGMLTFSGPKVQSGLDAGHSRAIP  
 VSREKFPSSAPSS"  
 535..1050  
**3'UTR**  
**ORIGIN**

Alignment Scores:  
 Pred. No.: 2,996-58 Length: 1056  
 Score: 599.00 Matches: 115  
 Percent Similarity: 97.56% Conservative: 5  
 Best Local Similarity: 93.50% Mismatches: 3  
 Query Match: 93.74% Indels: 0  
 DB: 10 Gaps: 0  
 US-10-657-740-1\_COPY\_51\_173 (1-123) x RNU47922 (1-1056)

QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20  
 DB 163 TCTCTTCCGCACAGTGTGGACTCCGGCATCTCTGAGTCCGATCTGACCGGACAG 222  
 QY 21 PheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrValLysValGln 40  
 DB 223 TTTGTCACTTCTTGGATGTGAAGCACATCTCTCTGAGGACCTCACCGTGAAGTACTG 282  
 QY 41 AspAspPheValGluIleHisGlyLysHisGlnArgGlnAspAspHisGlyTyrIle 60  
 DB 283 GAAGATTTCGTGGAGATCATGGCAACACACACAGAGGACGAGTACATGGCTTACATT 342  
 QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
 DB 343 TCCCGTGAATTCACCGTCGCTACCGTCTCCATGTGGACCATGCGCCCTCTCC 402  
 QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100  
 DB 403 TGCTCTTGTCTCGGATGGCATGTGACCTTCTGCGCCCAAGTCCAGTCTGCGCTG 462  
 QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120  
 DB 463 GATGTGGCCACAGCAGAGGAGGCCATTCCCGTGTCCAGGGAGGAGAGCCAGCTCGGCA 522  
 QY 121 ProSerSer 123  
 DB 523 CCCTCGTCC 531

RESULT 9  
 BOVCRYA  
 LOCUS BOVCRYA 793 bp mRNA linear MAM 26-APR-1993  
 DEFINITION Bovine alpha-A-crystallin gene, complete cds.  
 ACCESSION M26142

M26142.1 GI:162909  
 crystallin.  
**SOURCE** Bos taurus (cow)  
**ORGANISM** Bos taurus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 Bovinae; Bos.  
**REFERENCE** 1 (bases 1 to 793)  
**AUTHORS** Hay, R.E. and Petrash, J.M.  
**TITLE** Nucleotide sequence of a bovine lens alpha A-crystallin cDNA  
**JOURNAL** Biochem. Biophys. Res. Commun. 148 (1), 31-37 (1987)  
**MEDLINE** 88049675  
**PUBMED** 3675580  
**COMMENT** Original source text: Bovine lens, cDNA to mRNA, clone  
 pBL-alpha-A2-1.  
**FEATURES**  
 Location/Qualifiers  
 1..793  
 /organism="Bos taurus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9913"  
 31..552  
 /note="alpha-A-crystallin"  
 /codon\_start=1  
 /protein\_id="AAA30471.1"  
 /db\_xref="GI:162910"  
 /translation="MDIAIQHPWFKRTLGPVPSRLFDQFGEGLFEYDLPLFLSSTI  
 SPYRQSLFRVLDSGISSEVSRDRDKFVILDKHFSPEDLTVKQEDFVEIHGKNE  
 RQDDHGYISREHFHRRYRLPSNVDSQALSCLSDGMLTFSGPKIPSGVDAGHSRAIP  
 VSREKFPSSAPSS"  
**CDS**  
 29 bp upstream of NcoI site.

Alignment Scores:  
 Pred. No.: 2,786-58 Length: 793  
 Score: 598.00 Matches: 115  
 Percent Similarity: 97.56% Conservative: 5  
 Best Local Similarity: 93.50% Mismatches: 3  
 Query Match: 93.58% Indels: 0  
 DB: 4 Gaps: 0  
 US-10-657-740-1\_COPY\_51\_173 (1-123) x BOVCRYA (1-793)

QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20  
 DB 181 TCCCTCTTCCGCACCGTGTGGACTCCGGCATCTCTGAGTCCGATCTGACCGGACAG 240  
 QY 21 PheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrValLysValGln 40  
 DB 241 TTTGTCACTTCTTGGATGTGAAGCACATCTCTCCGAGGACCTGACCGTGAAGTGTGAG 300  
 QY 41 AspAspPheValGluIleHisGlyLysHisGlnArgGlnAspAspHisGlyTyrIle 60  
 DB 301 GAGGACTTCGTGGAGATCCAGCGCAAGCACACAGCGGACGAGTACATGGCTTACATC 360  
 QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
 DB 361 TCCCGGAGTTCACCGCCGCTACCGCTCTCCAAACGTCGACGTCGCGACTCTCC 420  
 QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100  
 DB 421 TGCTCTTGTCTCGGATGGCATGTGACCTTCTGCGCCCAAGTCCATCTGCGGTG 480  
 QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120  
 DB 481 GACGCGGCCACAGCAGCGGCCCATCCCGTGTCCCGGAGGAGAGCCAGCTCTGCG 540  
 QY 121 ProSerSer 123  
 DB 541 CCCTCGTCC 549

RESULT 10  
 RNACRY  
 LOCUS RNACRY 885 bp mRNA linear ROD 08-JAN-1999  
 DEFINITION Partial sequence of messenger RNA for rat alpha-a2-crystallin.

ACCESSION V01219 J00715  
 VERSION V01219.1 GI:55598  
 KEYWORDS complementary DNA; crystallin.  
 SOURCE Rattus norvegicus (Norway rat)  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 REFERENCE 1 (bases 1 to 885)  
 AUTHORS Moormann, R.J., van der Velden, H.M., Dodemont, H.J., Andreoli, P.M.,  
 Bloemendal, H., and Schoenmakers, J.G.  
 TITLE An unusually long non-coding region in rat lens alpha-crystallin  
 messenger RNA  
 JOURNAL Nucleic Acids Res. 9 (19), 4813-4822 (1981)  
 MEDLINE 82081811  
 PUBMED 6171772  
 FEATURES  
 source  
 1..885 Location/Qualifiers  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10116"  
 <1..885  
 <1..368  
 /codon\_start=3  
 /product="crystallin"  
 /protein\_id="CAA24530.1"  
 /db\_xref="GI:809074"  
 /translation="PRTVLDGSGISEVRSDRDKFVFLDVKHPSPEDLTVKVLDFVEI  
 HGKNERQDDHGYSIRFHRRLPSNVQDALSCLSDAGMLTFPGPKVQSGLDGHS  
 SERAIPVSRREKPSAPSS"  
 885  
 polyA\_site  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1,978-57 Length: 885  
 Score: 591.00 Matches: 113  
 Percent Similarity: 97.52% Conservative: 5  
 Best Local Similarity: 93.33% Mismatches: 3  
 Query Match: 92.43% Indels: 0  
 DB: 10 Gaps: 0  
 US-10-657-740-1\_COPY\_51\_173 (1-123) x RNACRY (1-885)  
 QY 3 PheArgThrValLeuAspSerGlyIleSerGluValArgSerAspAspAspLysPheVal 22  
 DB 3 TTCCGCACAGTGTGGACTCCGGCATCTCTGAGGTCGGATCTGACCGGGACAGTTTGTTC 62  
 QY 23 IlePheLeuAspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAsp 42  
 DB 63 ATCTTCTTGGATGTGAAGCACCTTCTCTCTGAGGACCTCACCGTGAAGTACTGGAAGAT 122  
 QY 43 PheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArg 62  
 DB 123 TTCGTGGAGATCCATGGCAACACACACGAGAGGAGGATGACCATGGCTTACATTTCCCGT 182  
 QY 63 GluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSer 82  
 DB 183 GAATTTACCGTCGCTACCGTCTGCTTCCATGTGGACCATCGCCCTCTCTGCTCC 242  
 QY 83 LeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAla 102  
 DB 243 TTGCTGGGATGGCATGTGACCTTCTCTGGCCCAAGGTCCAGTCTGGCTTGGATGCT 302  
 QY 103 ThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAlaProSer 122  
 DB 303 GCCACACGAGAGGGCCATTCCCGTGTACACGGAGGAGAACCCAGCTCGGCACCCCTCG 362  
 QY 123 Ser 123  
 DB 363 TCC 365  
 RESULT 11  
 MMU310308

LOCUS MMU310308 796 bp mRNA linear ROD 03-APR-2001  
 DEFINITION Mus musculus mRNA for alpha-A-crystallin (Cryaa gene).  
 ACCESSION AJ310308  
 VERSION AJ310308.1 GI:13548627  
 KEYWORDS alpha-A-crystallin; cryaa gene.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1  
 AUTHORS Graw, J.  
 TITLE Characterization of a new, dominant V124E mutation in the mouse  
 alpha-A-crystallin encoding gene  
 JOURNAL Unpublished  
 AUTHORS 2 (bases 1 to 796)  
 TITLE Direct Submission  
 JOURNAL Submitted (02-APR-2001) Graw J., Institute of Mammalian Genetics,  
 GSF-National Research Center for Environment and Health,  
 Ingolstaedter Landstr. 1, Neuherberg, Germany, D-85764, GERMANY  
 FEATURES  
 source  
 1..796 Location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C3H/El"  
 /db\_xref="taxon:10090"  
 /tissue\_type="Lens"  
 1..796  
 /gene="Cryaa"  
 52..642  
 /gene="Cryaa"  
 /function="lens structural protein; chaperone"  
 /note="alternative splicing"  
 /codon\_start=1  
 /product="alpha-A-crystallin"  
 /protein\_id="CAC35974.1"  
 /db\_xref="GI:13548628"  
 /translation="MDVTIOHPWPKRALPPYPSRLFDORFEGLEFYEDLLPFLSSTI  
 SPYRQSLFRVLDSGISELTMWFMVHOPHAGNKNPNVKSRSRDRDKFVFLDVKH  
 FSPEDLTVKVLDFVEIHGKNERQDDHGYSIRFHRRLPSNVQDALSCLSDAG  
 MLTFSGPKVQSGLDGHSRAIPVSRREKPSAPSS"  
 241..309  
 /gene="Cryaa"  
 /note="alternative splicing"  
 exon  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 5,918-56 Length: 796  
 Score: 577.50 Matches: 115  
 Percent Similarity: 82.19% Conservative: 5  
 Best Local Similarity: 78.77% Mismatches: 3  
 Query Match: 90.38% Indels: 23  
 DB: 10 Gaps: 1  
 US-10-657-740-1\_COPY\_51\_173 (1-123) x MMU310308 (1-796)  
 QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGlu-----ValArgSerAsp 17  
 DB 202 TCCCTCTTCCGACTGTGCTGGACTCGGGCATCTCTGAGCTCATGACCATATGTGGTTT 261  
 QY 14 -----ValArgSerAsp 17  
 DB 262 GTAATGCACCAACACATGCTGGAAACCCCAAGAACCCCGTCAAGTCCGATCTGAC 321  
 QY 18 ArgAspLysPheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrVal 37  
 DB 322 CGGACCAAGTTGTTCATCTTCTGACGTGAAGCATTCTCTCTGAGGACCTCACCGTG 381  
 QY 38 LysValGlnAspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHis 57  
 DB 382 AAGTACTGGAGGATTTTGTGAGATTACCGCAACACACACGAGAGCAGATGACCAT 441  
 QY 58 GlyTyrIleSerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSer 77



```

Alignment Scores:
Pred. No.: 3.97e-53 Length: 447
Score: 550.00 Matches: 103
Percent Similarity: 99.09% Conservative: 6
Best Local Similarity: 93.64% Mismatches: 1
Query Match: 86.07% Indels: 0
DB: 4 Gaps: 0

US-10-657-740-1_COPY_51_173 (1-123) x AJ617725 (1-447)

QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20
Db 118 TCCTCTTCGCGACCGTCTCGACTCCGGCATCTCCGAGGTCCGGTCTGATCGGGACCAAG 177
QY 21 PheValIlePheLeuAspValIleHisPheSerProGluAspLeuThrValIleValGln 40
Db 178 TTCCTCATCTCTGACGTGAGCACTTCTCCCTGAGGACCTGACTGTGAAGGTGCAG 237
QY 41 AspAspPheValGluIleHisGlyIleHisGlyIleHisGlnAspAspHisGlyIle 60
Db 238 GATGACTTTGTGGAGATCCATGGCAACCAATGAGAGCGAGGACGACCGCTACATC 297
QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80
Db 298 TCCCGTGAGTTCACCGCGCTACCGCTGCCCTCCAACTGAGGACCGACCTGCACTCTCT 357
QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100
Db 358 TGCTCCCTCTCTCGGACCGGATGCTGACCTTCTGTGGCCCCCAAGATCCAGTCTGGCATG 417
QY 101 AspAlaThrHisAlaGluArgAlaIlePro 110
Db 418 GATGCCAGGCACAGTGAGAGGCCATCCCC 447

RESULT 14
AJ617726 477 bp mRNA linear VRT 15-JUL-2004
LOCUS Sphenodon punctatus partial mRNA for alphaA-crystallin (cryaa
DEFINITION
ACCESSION AJ617726
VERSION AJ617726.1 GI:50344348
KEYWORDS alphaA-crystallin; cryaa gene.
SOURCE Sphenodon punctatus (tuatara)
ORGANISM Sphenodon punctatus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Lepidosauria; Sphenodontia; Sphenodontidae; Sphenodon.
TITLE van Rheede, T. and de Jong, W.W.
JOURNAL The alpha-crystallins of the platypus Ornithorhynchus anatinus:
REFERENCE Origin of the alternatively spliced exon alphaAins and implications
AUTHORS for mammalian phylogeny
TITLE Unpublished
JOURNAL 2 (bases 1 to 477)
REFERENCE Franck, E.
AUTHORS Direct Submission
JOURNAL Submitted (15-DEC-2003) Franck E., Biochemistry, Radboud University
TITLE Niimegen, 9101, 6500 HB Niimegen, NETHERLANDS
JOURNAL Location/Qualifiers
FEATURES
source
1..477
/organism="Sphenodon punctatus"
/mol_type="mRNA"
/db_xref="taxon:8508"
1..477
/gene="cryaa"
<1..477
/gene="cryaa"
/codon_start=1
/product="alphaA-crystallin"
/db_xref="GI:50344349"
/translaton="PLIPSRLLFDFFGEGLEPDLPLFGSTNIPYRQGFRTVLRS
GISEVSRDRDKFTIFLDVHKFSPEDLSVKIIDDVFEVHKHNERQVDHGVISREFFHR
YRLPSNVDSQSAITCSLSDGDMWTFSAKPVQSNMDPSHSERPVPVSREKFTSAPSS"

ORIGIN
Alignment Scores:
Pred. No.: 7.22e-53 Length: 477
Score: 548.00 Matches: 102
Percent Similarity: 91.87% Conservative: 11
Best Local Similarity: 82.93% Mismatches: 10
Query Match: 85.76% Indels: 0
DB: 5 Gaps: 0

US-10-657-740-1_COPY_51_173 (1-123) x AJ617726 (1-477)

QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20
Db 106 TCCTTTTCCGCGACCGTCTCGAATCAGGCATTTTCAGAGGTGAGATCTGACCGGACAAG 165
QY 21 PheValIlePheLeuAspValIleHisPheSerProGluAspLeuThrValIleValGln 40
Db 166 TTTCAATCTTCTCGGATGTAAACACTTCTCTCTGAGATTTCAGTGTGAAGATTATT 225
QY 41 AspAspPheValGluIleHisGlyIleHisGlnAspAspHisGlyIle 60
Db 226 GATGACTTTGTGGAATCCACGCAAGCACAAATGAGAGACAGTAGACCATGGCTACATC 285
QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80
Db 286 TCCCGGGAATTCACCGCGAGTACCGCTGCCCTTCCAAATGTGGACCAATCCGCCATCATC 345
QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100
Db 346 TGCTCCCTCTCTGTGTGAGCGCATGATGACCTTCTCTGCCCCCAAGGTCCAGTCTAACATG 405
QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120
Db 406 GACCCGACGACGAGAGAGACCCATTCCTGTATCCGTGGAAGAGAAGCCACCTCGCGCG 465
QY 121 ProSerSer 123
Db 466 CCTTCTCTCC 474

RESULT 15
RCAACRYST 522 bp mRNA linear VRT 18-SEP-1995
LOCUS R.catesbeiana mRNA for alpha-crystallin alpha A subunit.
DEFINITION
ACCESSION X85205
VERSION X85205.1 GI:732915
KEYWORDS aa-crystallin.
SOURCE Rana catesbeiana (bullfrog)
ORGANISM Rana catesbeiana
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Amphibia; Batrachia; Anura; Neobatrachia; Ranioidea; Rana;
TITLE Aquarana.
JOURNAL 1
AUTHORS Lu, S.F., Pan, F.M. and Chiou, S.H.
TITLE Sequence analysis of frog alpha-crystallin cDNA and its deduced
REFERENCE primary structure: comparison of alpha A subunit chains among
AUTHORS different vertebrate species
JOURNAL Biochem. Biophys. Res. Commun. 210 (3), 974-981 (1995)
MEDLINE 95283564
PUBMED 7763271
REFERENCE 2 (bases 1 to 522)
AUTHORS Chiou, S.H.
TITLE Direct Submission
JOURNAL Submitted (10-MAR-1995) S.H. Chiou, Institute of Biological
REFERENCE Chemistry, Academia Sinica, Institute of Biochemical Sciences,
AUTHORS National Taiwan University, P.O. Box 23-106, Taipei 10098, TAIWAN
TITLE Sequence overlapping with that under the acc#X00716.
FEATURES
source
1..522
/organism="Rana catesbeiana"
/mol_type="mRNA"
/db_xref="taxon:8400"
1..522

```



Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Neobatrachia; Ranoidea; Ranidae; Rana;  
Rana.

# REFERENCE

Tomarev, S.I., Zinov'eva, R.D., Dolgilevich, S.M., Kravay, A.S.,  
Skraybin, K.G. and Gause, G.G. Jr.

The absence of the long 3'-non-translated region in mRNA coding for  
eye lens alpha A2-crystallin of the frog (*Rana temporaria*)

FEBs Lett. 162 (1), 47-51 (1983)

84005173

6604666

2 (bases 1 to 765)

Tomarev, S.I., Zinov'eva, R.D., Kraev, A.S., Skraybin, K.G. and  
Gause, G.G.

Primary structure of cloned cDNA coding alpha-A2 crystallin of the  
eye lens of the frog *Rana temporaria*

Dokl. Biochem. 271, 277-280 (1984)

Data kindly reviewed (05-11-1985) by S.I. Tomarev

Sequence 1 to 185 is complementary to sequence 553 to 369 and  
probably a cloning artefact; amino terminal amino acids are  
missing.

# FEATURES

source Location/Qualifiers

1..765 /organism="Rana temporaria"

/mol\_type="mRNA"

/db\_xref="taxon:8407"

1..185 /note="artefactual sequence"

186..632 /note="unnamed protein product; Protein sequence is in  
conflict with the conceptual translation; alpha-A2  
crystalline (aa 25 to 173)"

/codon\_start=1

/protein\_id="CAA25308.1"

/db\_xref="GI:1334774"

/db\_xref="GO:PO2508"

/translation="QVFGEGMFDYDLFPFLTSTVSPHYRHGLRGFMDSGISEVRSR  
DFTNLDVKHPSDDLTWKILDVFEIHKHSRQDDHGVISREHRYRLPNLQD  
SSISGSLSDAGILTFSGPKMNSLVSHSERPIVSRREKETSAPSS"

744..749

/note="polyadenylation signal"

765

/note="polyadenylation site"

# ORIGIN

Alignment Scores:

Pred. No.: 2 45e-49 Length: 765  
Score: 519.00 Matches: 97  
Percent Similarity: 90.98% Conservative: 14  
Best Local Similarity: 79.51% Mismatches: 11  
Query Match: 81.22% Indels: 0  
DB: 5 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x RTCRYA (1-765)

QY 2 LeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPhe 21

Db 267 CTCCTCCGAGGATTCATGCATCCGGCATCTCTGAGTGCGTTCAGATCCGATCGTTC 326

QY 22 ValIlePheLeuAspValIleHisPheSerProGluAspLeuThrValIysValGlnAsp 41

Db 327 ACCATTACTTGGAGCTGAGCATTTCTCCCGCATGATCTGATGTGAAGATTCTGGAT 386

QY 42 AspPheValGluIleHisGlyIleHisAsnGluArgGlnAspAspHisGlyTyrIleSer 61

Db 387 GACTTTGTGGAATCATGGAAGACACAGCAGAGGAGGATGACCATGCTCATATCC 446

QY 62 ArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCys 81

Db 447 CGGAGTGTCCACCGCGCTACCGCTCCCTCCCAACCTGGACCAATCTCTATCAGCTGC 506

QY 82 SerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAsp 101

Db 507 TCACCTCTCGCGATTCGACCTTCAGCGACCAAAATGATGTCAAACCTGTCTC 566

QY 102 AlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAlaPro 121

Db 567 TCACGCCACAGCAGGAGGCCATCCCTGTGTCCAGAGAGGAAAGCCACCTCTGCCCCC 626

QY 122 SerSer 123

Db 627 TCCTCC 632

RESULT 18

D86299

LOCUS

DEFINITION

ACCESION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

CDS

# ORIGIN

Alignment Scores:

Pred. No.: 1 11e-47 Length: 1361  
Score: 507.00 Matches: 93  
Percent Similarity: 89.43% Conservative: 17  
Best Local Similarity: 75.61% Mismatches: 13  
Query Match: 79.34% Indels: 0  
DB: 5 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x D86299 (1-1361)

QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20

Db 292 TCATGTTCCGACACTACCTGACCTCTCGCATATCCGAGGTGAGTCTCTGCGCGACAG 351

QY 21 PheValIlePheLeuAspValIleHisPheSerProGluAspLeuThrValIysValGln 40

Db 352 TTCAGATTACCTGGAGCTCAAGCACTCTCACCGGAAGACCTAAGTGTCAAGATCCTC 411

QY 41 AspAspPheValGluIleHisGlyIleHisAsnGluArgGlnAspAspHisGlyTyrIle 60



```

Db      412 GACGATTTACGTTGGAGATCATGGAAGACACAGCGACCGGAGATCATCGGATATGTC 471
Qy      61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80
Db      472 TCGGAGAGATTCCACCGCGCTATCGCTCGCGCGAGCGTGGACAGTCTCCATCACA 531
Qy      81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrClyLeu 100
Db      532 TGCTCCCTGTCACCTGATGGGATGCTGACCTTCTCGGGCACAAGATGATGCTAGCTTG 591
Qy      101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120
Db      592 GACTCCACCGCGGAGAGGCCCATCTCTGTTGCGAGAGAGAGAGGCCCGCTCTGCC 651
Qy      121 ProSerSer 123
Db      652 CAATCCTCT 660

RESULT 19
AJ617727
LOCUS   Lygodactylus picturatus partial mRNA for alphaA-crystallin (cryaA
DEFINITION
ACCESSION AJ617727
VERSION   AJ617727.1 GI:50344350
KEYWORDS  alphaA-crystallin; cryaA gene.
SOURCE    Lygodactylus picturatus
ORGANISM  Lygodactylus picturatus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Lepidosauria; Squamata; Scleroglossa; Gekkota; Gekkonidae;
Lygodactylus.
REFERENCE 1
AUTHORS   van Rheede, T. and de Jong, W.W.
TITLE     The alpha-crystallins of the platypus Ornithorhynchus anatinus:
          Origin of the alternatively spliced exon alphaAins and implications
          for mammalian phylogeny
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 447)
AUTHORS   Franck, E.
TITLE     Direct Submission
JOURNAL   Submitted (16-DEC-2003) Franck E., Biochemistry, Radboud University
          Nijmegen, 9101, 6500 HB Nijmegen, NETHERLANDS
FEATURES  Location/Qualifiers
          source
            1..447
            /organism="Lygodactylus picturatus"
            /mol_type="mRNA"
            /db_xref="taxon:57064"
            1..447
            /gene="cryaA"
            /gene="cryaA"
            /codon_start=1
            /product="alphaA-crystallin"
            /protein_id="CAF02103.1"
            /db_xref="GI:50344351"
            /translation="RAFGPLPSPRLFDQFFGGLLYDYLPLFSTISPPYRQSLFRT
          VLESGVSEVRSDRDKFTFLDKVHFSPEDLSVKVDDFVEIHGKHNERQDDHGYSIRE
          FHRRYLPSNVDSQSAISCSLSADGMLTFAPKVSQNTDLSHERPIP"

Alignment Scores:
Pred. No.: 5.07e-48 Length: 447
Score: 505.00 Matches: 94
Percent Similarity: 92.73% Conservative: 8
Best Local Similarity: 85.45% Mismatches: 8
Query Match: 79.03% Indels: 0
DB: 5 Gaps: 0

US-10-657-740-1_COPY_51_173 (1-123) x AJ617727 (1-447)

Qy      1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20

```

```

Db      118 TCTCTCTCCCACTGTTCTTGAGTCAGGTGTTCTTGAGGTGAGATCTGACGGGACAAG 177
Qy      21 PheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrValLysValGln 40
Db      178 TTTACAATCTTTTGGATGTAAACATTTCTCACCTGAAGATTTGAGCGTGAAGGTGATC 237
Qy      41 AspAspPheValGluIleHisGlyLysHisAlaGluArgGlnAspAspHisGlyTyrIle 60
Db      238 GATGACTTTTGGAAATCCATGGCAAAACACAAATGAGACAGGATGATCATGGGTACATC 297
Qy      61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80
Db      298 TCCGCTGAATCCACCGCAGGTACCGCTCTCCATGTGGACCATGCGCCATATCC 357
Qy      81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100
Db      358 TGCTCTCTGCTGCTGATGGCATGCTGACCTTCGCTGCCCTTAAGTCCAGTCCCAACT 417
Qy      101 AspAlaThrHisAlaGluArgAlaIlePro 110
Db      418 GACCTTAGCCACAGTACGAGACCCATTCTCT 447

RESULT 20
TSU31938
LOCUS   Trachemys scripta elegans alpha A-crystallin mRNA, partial cds.
DEFINITION
ACCESSION U31938
VERSION   U31938.1 GI:1223846
KEYWORDS
SOURCE    Trachemys scripta elegans
ORGANISM  Trachemys scripta elegans
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Testudines; Cryptodira; Testudinoidea; Emydidae; Trachemys.
REFERENCE 1 (bases 301 to 446)
AUTHORS   Hedges, S.B., Simmons, M.D., van Dijk, M.A., Caspers, G.J., de
          Jong, W.W. and Sibley, C.G.
TITLE     Phylogenetic relationships of the hoatzin, an enigmatic South
          American bird
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 92 (25), 11662-11665 (1995)
MEDLINE   96102174
PUBMED    8524824
REFERENCE 2 (bases 1 to 448)
AUTHORS   Caspers, G.J., Reinders, G.J., Leunissen, J.A., Wattel, J. and de
          Jong, W.W.
TITLE     Protein sequences indicate that turtles branched off from the
          amniote tree after mammals
JOURNAL   J. Mol. Evol. 42 (5), 580-586 (1996)
MEDLINE   96259313
PUBMED    8662010
REFERENCE 3 (bases 1 to 448)
AUTHORS   de Jong, W.W.
TITLE     Direct Submission
JOURNAL   Submitted (20-JUL-1995) Wilfried W. de Jong, Biochemistry,
          University of Nijmegen, P.O.Box 9101, 6500 HB Nijmegen, The
          Netherlands
FEATURES  Location/Qualifiers
          source
            1..448
            /organism="Trachemys scripta elegans"
            /mol_type="mRNA"
            /sub_species="elegans"
            /db_xref="taxon:31138"
            /tissue_type="lens"
            <1..448
            /note="related to small heat shock proteins"
            /codon_start=2
            /product="alpha A-crystallin"
            /protein_id="AAB08829.1"
            /db_xref="GI:1223847"
            /translation="RALGPLPSPRLPDQYLGELFDYDYLPLFPSTISPPYRHSILFRT
          VLESGVSEVRSDRDKFTFLDKVHFSPEDLSVKIMDDFVEIHGKHNERQDDHGYSIRE
          FHRRYLPSNVDSQSAITCSLSADGMLTFSGKVSQNMDSYSERP"

CDS

```

ORIGIN

Alignment Scores:  
 Pred. No.: 1,11e-47 Length: 448  
 Score: 502.00 Matches: 92  
 Percent Similarity: 93.64% Conservative: 11  
 Best Local Similarity: 83.64% Mismatches: 7  
 Query Match: 78.56% Indels: 0  
 DB: 5 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x TSU31938 (1-448)

QY 1 SerLeuPheArgThrValLeuAspSerGlyHisValSerGluValArgSerAspArgAspLys 20  
 Db 119 TCCTCTTCGCGACCGTCTCGGAATCAGGATTCAGAGGTGAGTCTGACCGGACAAAG 178  
 QY 21 PheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrValLysValGln 40  
 Db 179 TTTTACAATCTCTGGATGTAAACACTTCTCCCGAAGATCTGAGTGTGAAGATTATG 238  
 QY 41 AspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIle 60  
 Db 239 GATGACTTTGTGAAATCCATGTCGAAGCACAATGAGACAGCAGCATGGTACATT 298  
 QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
 Db 299 TCCCGGAATTCACCGCAGATACCGCTGCGCTTCCACGTGGACCAATCTGCCATCAC 358  
 QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100  
 Db 359 TGCTCCTCTGCTGATGCGATGCTGACTTCTCTGGCCCAAAAGTCCAGTCCCAATG 418  
 QY 101 AspAlaThrHisAlaGluArgAlaIlePro 110  
 Db 419 GACACCAGTATAGCGAGAGCCCATTCCT 448

RESULT 21  
 D88185  
 LOCUS  
 DEFINITION  
 XENOPUS LAEVIS mRNA FOR ALPHA A CRYSTALLIN, COMPLETE CDS.  
 ACCESSION  
 D88185.1 GI:4589827  
 VERSION  
 alpha A crystallin.  
 KEYWORDS  
 XENOPUS LAEVIS (AFRICAN CLAWED FROG)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
 Xenopodinae; Xenopus; Xenopus.  
 REFERENCE  
 Mizuno, N., Mochii, M., Takahashi, T. C., Ezuchi, G. and Okada, T. S.  
 Lens regeneration in Xenopus is not a mere repeat of lens  
 development, with respect to crystallin gene expression  
 Differentiation 64 (3), 143-149 (1999)  
 JOURNAL  
 MEDLINE  
 PUBMED  
 99251119  
 10234811  
 2 (bases 1 to 705)  
 AUTHOR  
 Mizuno, N.  
 TITLE  
 Direct Submission  
 JOURNAL  
 Submitted (30-SEP-1996) Nobuhiko Mizuno, Biohistory Research Hall,  
 Laboratory Div., Murasaki Chou 1-1, Takatuki, Osaka 569-11, Japan  
 (E-mail: Nobuhiko.Mizuno@ate.brh.co.jp, Tel: 0726-81-9751,  
 Fax: 0726-81-9757)  
 COMMENT  
 Sequence updated (12-Apr-1999).  
 FEATURES  
 Location/Qualifiers  
 1..705  
 /organism="Xenopus laevis"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:8355"  
 67..582  
 /codon\_start=1  
 /product="alpha A crystallin"  
 /protein\_id="BAA76897.1"  
 /db\_xref="GI:4589828"  
 CDS

RODDHGYSREFHRRYRLPSNMDQNSVSTLSADGILLTFQPKLQSNMDSHSDRTIP  
 VSKEKSGSSS"  
 705  
 /note="22 a nucleotides"

polyA\_site  
 ORIGIN

Alignment Scores:  
 Pred. No.: 2,58e-46 Length: 705  
 Score: 492.00 Matches: 91  
 Percent Similarity: 90.83% Conservative: 18  
 Best Local Similarity: 75.83% Mismatches: 11  
 Query Match: 77.00% Indels: 0  
 DB: 5 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x D88185 (1-705)

QY 1 SerLeuPheArgThrValLeuAspSerGlyHisValSerGluValArgSerAspArgAspLys 20  
 Db 217 AACCTCTCCAGAGATACCTGGACTCTGGGATCTCTGAGGTGCGTTGAGACCGTGACCGT 276  
 QY 21 PheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrValLysValGln 40  
 Db 277 TTTGTCAATTACCTGGATGTGAAGCATTTCTCCCTGAAGATCTGAGCGTCAAAAGTCCAT 336  
 QY 41 AspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIle 60  
 Db 337 GATGACTTTGTAGAGATTGAGGAAACACACAGCAGCAGGATGATCAGGATACATC 396  
 QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
 Db 397 TCCCGAATTCATCGCGATATCGTCTTCTTCCATATGATGATGAGTCTGTGAGC 456  
 QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100  
 Db 457 TGCACTCTGTCTCGGACGGATCTCTCTTATTCGTCTCCCAACTGCAATCCACATG 516  
 QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120  
 Db 517 GACTCCAGCCACAGCAGATAGGACCATCTCTGTGTCCAGGAGAGAAATCAGGCTCATCC 576

RESULT 22  
 EUDLAACRYS  
 LOCUS  
 DEFINITION  
 EUDROMIA ELEGANS ALPHA A-CRYSTALLIN mRNA, PARTIAL CDS.  
 ACCESSION  
 L25850  
 VERSION  
 L25850.1 GI:538348  
 KEYWORDS  
 alpha A-crystallin.  
 SOURCE  
 Eudromia elegans (elegant crested-tinamou)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Palaeognathae; Tinamiformes; Tinamidae;  
 Eudromia.  
 REFERENCE  
 1 (bases 1 to 448)  
 Caspers, G. J., Wattel, J. and de Jong, W. W.  
 TITLE  
 Alpha A-crystallin sequences group tinamou with ratites  
 JOURNAL  
 Mol. Biol. Evol. 11 (4), 711-713 (1994)  
 MEDLINE  
 94359400  
 PUBMED  
 8078410  
 COMMENT  
 Original source text: Eudromia elegans eye lens cDNA to mRNA.  
 FEATURES  
 Location/Qualifiers  
 1..448  
 /organism="Eudromia elegans"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:8805"  
 /tissue\_type="eye lens"  
 <1..448  
 /codon\_start=2  
 /product="alpha A-crystallin"  
 /protein\_id="AAA49254.1"  
 /db\_xref="GI:538349"  
 CDS

/translation="BALGPLPSRLFDQFGGLLEYDLLPLFSSITSPYKQSLFRS  
 VLESGLSVRSRDEKFTMLDGHFSPDLVKIIDDVFEIHKHSERQDDHGYSIRE  
 FHRRYLSNVDSQSAITCSLSDDGMLTFSGGPKVQANMDPSHSERP"

## ORIGIN

Alignment Scores:  
 Pred. No.: 5.6e-46 Length: 448  
 Score: 487.00 Matches: 89  
 Percent Similarity: 92.73% Conservatives: 13  
 Best Local Similarity: 80.91% Mismatches: 8  
 Query Match: 76.21% Indels: 0  
 DB: 5 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x EUDLAACRYS (1-448)

QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20  
 DB 119 TCTCTCTCCGAGCGTCTGGAGTCAGCATTTTCAGAGGTGAGATCTGACCGGAAAAA 178  
 QY 21 PheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrValLysValGln 40  
 DB 179 TTTACAATCATGCTGGATGTAACACATCTCTCTGAAGACCTAAGCGTGAAGATCAT 238  
 QY 41 AspAspPheValGluIleHisGlyLysHisGlnArgGlnAspAspHisGlyTyrIle 60  
 DB 239 GATGACTTTGTGAAATCATGCGAAGCACAGTGAAGACAGACGACCATCGCTACATC 298  
 QY 61 SerArgGluPheHisArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
 DB 299 TCCCGTGAAATTCACCGCGGTACCGCTCCCAACGTGGACAGCTGCGCATCACC 358  
 QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100  
 DB 359 TGCTCCCTCTCAGTGACGGCATGTGACCTTCTCCGGCCCCNAGGTCAGCGCAACATG 418  
 QY 101 AspAlaThrHisAlaGluArgAlaIlePro 110  
 DB 419 GACCCAGCACAGGAGAGACCCATCCCT 448

## RESULT 23

APRAAC 448 bp mRNA linear VRT 09-SEP-2004  
 LOCUS A.platyrrhynchos mRNA for alpha-A-crystallin.  
 DEFINITION X96592  
 ACCESSION X96592.1 GI:1945629  
 VERSION alpha-A-crystallin.  
 KEYWORDS Anas platyrhynchos  
 SOURCE Anas platyrhynchos  
 ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Anseriformes; Anatidae; Anas.  
 REFERENCE 1  
 AUTHORS Caspers,G.J., Uit de Weerd,D., Wattel,J. and de Jong,W.W.  
 TITLE alpha-Crystallin sequences support a galliform/anseriform clade  
 JOURNAL Mol. Phylogenet. Evol. 7 (2), 185-188 (1997)  
 MEDLINE 97271643  
 PUBMED 9126559  
 REFERENCE 2 (bases 1 to 448)  
 AUTHORS Caspers,G.J.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-MAR-1996) G.J. Caspers, University of Nijmegen,  
 Department of Biochemistry, P.O.Box 9101, NL- 6500 HB Nijmegen,  
 NETHERLANDS

## FEATURES

source Location/Qualifiers  
 1..448  
 /organism="Anas platyrhynchos"  
 /mol\_type="mRNA"  
 /strain="Peking"  
 /db\_xref="taxon:8839"  
 /tissue\_type="lens"  
 <1..>448  
 /codon\_start=2  
 /product="alpha-A-crystallin"  
 /protein\_id="CAA65410.1"  
 /db\_xref="GI:1945630"  
 /db\_xref="GOA:O12984"  
 /db\_xref="UniProt/Swiss-Prot:O12984"

## CDS

/translation="RALCPLIPSLRFDQFGEGLLEYDLLPLFSTISPYRQSLFRS  
 VLESIGSEVRSDRDKPTIMLDVKKFSPEDLSVKIIDDVEIHGRKSRQDDHGYISRE  
 FHRRYLPANVDQSAITCSLSGDGMLTFSGPKVPSNMDPTHSERP"

## ORIGIN

Alignment Scores:  
 Pred. No.: 7.27e-46 Length: 448  
 Score: 486.00 Matches: 89  
 Percent Similarity: 91.82% Conservatives: 12  
 Best Local Similarity: 80.91% Mismatches: 9  
 Query Match: 76.06% Indels: 0  
 DB: 5 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x APRAAC (1-448)

QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20  
 DB 119 TCCCTCTCCGAGCGTCTGGAGTCGGCATTTTCAGAGGTGAGCTCTGACCGGACAAG 178  
 QY 21 PheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrValLysValGln 40  
 DB 179 TTTAGATCATGCTGGATGTAACACATCTCTCTGAAGATCTGAGCGTGAAGATATC 238  
 QY 41 AspAspPheValGluIleHisGlyLysHisGlnArgGlnAspAspHisGlyTyrIle 60  
 DB 239 GATGACTTTGTGAAATCATGCGAAGCACAGTGAAGACAGGATGACCGGCTACATC 298  
 QY 61 SerArgGluPheHisArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
 DB 299 TCCCGTGAGTTTACCGCGGTACCGCTCCCGCAACGTGGACAGCTGCGCATCACC 358  
 QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100  
 DB 359 TGCTCCCTCTCCGCGCAGCGCATGTGACCTTCTCAGGCCCCAGGTCCTCCCAACATG 418  
 QY 101 AspAlaThrHisAlaGluArgAlaIlePro 110  
 DB 419 GACCCACCCACAGCAGAGGCCCATCCCT 448

## RESULT 24

CLRAAC 448 bp mRNA linear VRT 09-SEP-2004  
 LOCUS C.livia mRNA for alpha-A-crystallin.  
 DEFINITION X96593  
 ACCESSION X96593.1 GI:1945730  
 VERSION alpha-A-crystallin.  
 KEYWORDS C.livia (domestic pigeon)  
 SOURCE C.livia  
 ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Columbiformes; Columbidae; Columba.  
 REFERENCE 1  
 AUTHORS Caspers,G.J., Uit de Weerd,D., Wattel,J. and de Jong,W.W.  
 TITLE alpha-Crystallin sequences support a galliform/anseriform clade  
 JOURNAL Mol. Phylogenet. Evol. 7 (2), 185-188 (1997)  
 MEDLINE 97271643  
 PUBMED 9126559  
 REFERENCE 2 (bases 1 to 448)  
 AUTHORS Caspers,G.J.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-MAR-1996) G.J. Caspers, University of Nijmegen,  
 Department of Biochemistry, P.O.Box 9101, NL- 6500 HB Nijmegen,  
 NETHERLANDS

## FEATURES

source Location/Qualifiers  
 1..448  
 /organism="Columba livia"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:8932"  
 /tissue\_type="lens"  
 <1..>448  
 /codon\_start=2  
 /product="alpha-A-crystallin"  
 /protein\_id="CAA65411.1"  
 /db\_xref="GI:1945731"

```

/db_xref="GOA:O12988"
/db_xref="UniProt/Swiss-Prot:O12988"
/translation="RALGPLIPSLRLEDFQFEGLELVLLPWFSSITSPYRSLFRS
VLESGISEVRSDREKFTMLDVKHSPEDLSVLIDDFVEIHGKHSERQDDHGYISRE
PHRYRLPANVDQAATCSLNDGMLTFSPKVPANNDASHGERPIP"
ORIGIN
Alignment Scores:
Pred. No.: 9,92e-45 Length: 448
Score: 476.00 Matches: 87
Percent Similarity: 90.91% Conservativeness: 13
Best Local Similarity: 79.09% Mismatches: 10
Query Match: 74.49% Indels: 0
DB: 5 Gaps: 0
US-10-657-740-1_COPY_51_173 (1-123) x CLRAAC (1-448)
QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspLys 20
Db 119 TCCCTCTTCGCGAGTGGAGTCAGGATTCAGAGTGAGGTCTGACCGGGAAG 178
QY 21 PheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrValLysValGln 40
Db 179 TTTACAATCATGCTGGATGTAAACACTTCTCCCGAAGACCTGAGTGTGAAGATTAT 238
QY 41 AspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIle 60
Db 239 GAGGACTTTGTGAAATCCATGTCAGACAGCAGTGAAGACAGCAGCAGCGTTACATC 298
QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80
Db 299 TCCCGGAAATTCACCGCGGTACCGCTGCGCGCCACGTCGACGAGTCCCATCAC 358
QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100
Db 359 TGTCTCCCTGTCACACGAGCATGCTGACCTTCTCGGCGCCCAAGTCCCGCCCAATG 418
QY 101 AspAlaThrHisAlaGluArgAlaIlePro 110
Db 419 GAGCGCAGCCAGCGGAGAGGCCATCCCC 448
RESULT 25
AY035778
LOCUS
DEFINITION
Danio rerio alpha A crystallin mRNA, complete cds.
ACCESSION
AY035778.1 GI:18266458
VERSION
1
KEYWORDS
SOURCE
Danio rerio (zebrafish)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE
1 (bases 1 to 730)
Runkle,S., Hill,J., Kantorow,M., Horwitz,J. and Posner,M.
Sequence and spatial expression of zebrafish (Danio rerio)
alphaA-crystallin
Mol. Vis. 8, 45-50 (2002)
JOURNAL
MEDLINE
21923047
PUBMED
11925526
REFERENCE
2 (bases 1 to 730)
Runkle,S., Hill,J., Kantorow,M., Horwitz,J. and Posner,M.
Direct Submission
Submitted (16-MAY-2001) Biology, Ashland University, 401 College
Avenue, Ashland, OH 44805, USA
LOCATION/Qualifiers
1. 730
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/tissue type="lens"
95..616
/note="small heat shock protein"
CDS
/codon_start=1
/product="alpha A crystallin"
/protein_id="AAK61363.1"
/db_xref="GI:18266458"
/translation="MDIAIQHPWFRTLYGYTRFLDQFFGEGLEFDYDLFPFTTSTVSP
YRSLFRNILDSSNGSVSRDREKFTVLDVKHSPEDLSVLIDDFVEIHGKHSERQDDHGYISRE
BRQDDHGYISREFHRRYRLPSNVDQSAITCTLSADGLITLCGPKTSGIDAGRGRTIP
VTREDKNSGSSS"
ORIGIN
Alignment Scores:
Pred. No.: 2,96e-44 Length: 730
Score: 474.00 Matches: 89
Percent Similarity: 84.92% Conservativeness: 18
Best Local Similarity: 70.63% Mismatches: 15
Query Match: 74.18% Indels: 4
DB: 5 Gaps: 2
US-10-657-740-1_COPY_51_173 (1-123) x AY035778 (1-730)
QY 1 SerLeuPheArgThrValLeuAsp-----SerGlyIleSerGluValArgSerAsp 17
Db 239 TCACTCTTCGCAACATCTGGACTCTCCAACTCGGAGTGTCTCTGAGGTGAGTCTGAC 298
QY 18 ArgAspLysPheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrVal 37
Db 299 AGAGAAATTTACAGTTTATCTGGATGTGAACATTTCTCTCTGATGAGTCAAGTGC 358
QY 38 LysValGlnAspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHis 57
Db 359 AAGGTGACAGATGACTATGTGGAGATCCAGGCGCAGCATCGAAGACAGGATGATCAT 418
QY 58 GlyTyrIleSerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSer 77
Db 419 GGTATACATCTCCGTCAGTTCCATCTGTCGTACCGCTGCTTCCAATGTGGACAGTCT 478
QY 78 AlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGln 97
Db 479 GCATACCTGACACTGCTGCTGATGGCTCTCACTTTTGTGGACCCCAAG---ACA 535
QY '98 ThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysPro 117
Db 536 AGTGGCATAGATCTGCGCGTGGAGATCGACCATCCCTGTTACCGCGGAGCAAGAGC 595
QY 118 ThrSerAlaProSerSer 123
Db 596 AACTCAGCTCTTCTCTCC 613
RESULT 26
BC083177
LOCUS
DEFINITION
Danio rerio cDNA clone MGC:92036 IMAGE:7045051, complete cds.
ACCESSION
BC083177
VERSION
BC083177.1 GI:52789216
KEYWORDS
SOURCE
MGC.
ORGANISM
Danio rerio (zebrafish)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE
1 (bases 1 to 1430)
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.P., Jordan,H., Moore,T., Max,A.I., Wang,J., Hong,L.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Frange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,

```

Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalilus, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932  
2 (bases 1 to 1430)  
Director MGC Project.  
Direct Submission  
Submitted (24-SEP-2004) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)  
Tissue Procurement: Len Zon, Harvard  
CDNA Library Preparation: Open Biosystems  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc\\_mgc@nhgri.nih.gov](mailto:nisc_mgc@nhgri.nih.gov)  
Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R., Maduro, Q.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Skantripp, S., Thomas, P.J., Touchman, J.W., Tsurgon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
Series: IRAP Plate: 178 Row: P Column: 11  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 23308654.  
passed the following selection criteria: matched mRNA gi: 23308654.

FEATURES  
source  
1..1430  
/organism="Danio rerio"  
/mol\_type="mRNA"  
/db\_xref="taxon:7955"  
/clone="MGC:92036 IMAGE:7045051"  
/tissue\_type="Whole, adult zebrafish"  
/clone\_lib="NIH ZGC\_10"  
/lab\_host="DH10B"  
/note="Vector: pExpress1"  
106..627  
/codon\_start=1  
/product="Unknown (protein for MGC:92036)"  
/protein\_id="AAH83177.1"  
/db\_xref="GI:52789217"  
/translation="MDIAIQHPWRRTGYTRLPDQFFGGLDLDLPFTTSTVSP  
YRSLFENILDSNGSVSEVRSDEKETVLDVKHFSPELRSVKVTDYDVEIQKHG  
ERQDDHGVISREFHRRYRLPSNVQSAITCTLSADGLLTLCGPKTSGIDAGRGRTIP  
VTREDKNSGSSS"

CDS  
6..5e-44 Length: 1430  
474.00 Matches: 89  
Percent Similarity: 84.92% Conservative: 18  
Best Local Similarity: 70.63% Mismatches: 15  
Query Match: 74.18% Indels: 4  
DB: 5 Gaps: 2

ORIGIN  
Alignment Scores:  
Pred. No.: 6.5e-44 Length: 1430  
Score: 474.00 Matches: 89  
Percent Similarity: 84.92% Conservative: 18  
Best Local Similarity: 70.63% Mismatches: 15  
Query Match: 74.18% Indels: 4  
DB: 5 Gaps: 2

US-10-657-740-1\_COPY\_51\_173 (1-123) x BC083177 (1-1430)

QY 1 SerLeuPheArgThrValLeuAsp-----SerGlyIleSerGluValArgSerAsp 17  
Db 250 TCACCTCTTCCCAACATCTCTGACTCTCCAACTCAGGTGCTCTGAGGTGAGGTCTGAC 309  
QY 18 ArgAspLysPheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrVal 37  
Db 310 AGAGAAATTTACAGTTTACCTGGATGTAACATTTCTCTCTGATGAGCTCAGTCTC 369  
QY 38 LysValGlnAspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHis 57  
Db 370 AAGGTGACAGATGACTATGTGAGATCCAGGCAAGCATGGAGAAAGACAGGATGATCAT 429  
QY 58 GlyTyrIleSerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSer 77  
Db 430 GGCTACATCTCCGTGAGTTCCATCGCGCTACCGCCCTCCCTTCAATGTGACCATGCT 489  
QY 78 AlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGln 97  
Db 490 GCCATCCTGCACACTGTCTGCTGATGCGCTGCTCACTCTTTGTGGACCAAG---ACA 546  
QY 98 ThrGlyLeuAspAlaThrHisAlaGluAlaGluAlaGluAlaGluAlaGluAlaGlu 117  
Db 547 AGTGCGATAGATGCTGGCGGTGGAGATCGACCATCCCTGTTACCCGCGAGCAAGAGC 606  
QY 118 ThrSerAlaProSerSer 123  
Db 607 AACTCGGGCTCTTCTCTCC 624

RESULT 27  
AY007972 676 bp mRNA linear VRT 23-OCT-2000  
LOCUS Clarias fuscus alpha-A crystallin mRNA, complete cds.  
DEFINITION AY007972  
ACCESSION AY007972  
VERSION AY007972.1 GI:10946518  
KEYWORDS  
SOURCE Clarias fuscus (whitespotted clarias)  
ORGANISM Clarias fuscus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;  
Clariidae; Clarias.  
REFERENCE  
1 (bases 1 to 676)  
AUTHORS Chlou, S.-H. and Yu, C.-M.  
TITLE Alpha crystallin of catfish eye lenses: cDNA and genomic analysis of alpha-A and alpha-B  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 676)  
AUTHORS Chlou, S.-H. and Yu, C.-M.  
TITLE Direct Submission  
JOURNAL Submitted (15-SEP-2000) Institute of Biological Chemistry, Academia Sinica, P.O. Box 23-106, Taipei, Taiwan 10098, Republic of China  
FEATURES  
source  
1..676  
/organism="Clarias fuscus"  
/mol\_type="mRNA"  
/db\_xref="taxon:33541"  
136..657  
/codon\_start=1  
/product="alpha-A crystallin"  
/protein\_id="AAG23866.1"  
/db\_xref="GI:10946519"  
/translation="MDIAIQHPWRRTLSNPSRLPDQFFGGLDLDLPFTTSTVSP  
YRSLFENILDSNGSVSEVRSDEKETVLDVKHFSPELRSVKVTDYDVEIQKHG  
ERQDDHGVISREFHRRYRLPSNVQSAITCTLSADGLLTLCGPKTSGYGRGRTIP  
VTREDDKNPAASS"

ORIGIN  
Alignment Scores:  
Pred. No.: 2.19e-43 Length: 676  
Score: 466.00 Matches: 89  
Percent Similarity: 84.13% Conservative: 17  
Best Local Similarity: 70.63% Mismatches: 16  
Query Match: 72.93% Indels: 4  
DB: 5 Gaps: 2

```

US-10-657-740-1_COPY_51_173 (1-123) x AY007972 (1-676)

QY 1 SerLeuPheArgThrValLeuAsp-----SerGlyIleSerGluValArgSerAsp 17
Db 280 TCTCTCTCCGAGATTTTGGACTCTCTCAACTCTGGCAATTCGAGGTGAGGTCTGAC 339

QY 18 ArgAspLysPheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrVal 37
Db 340 AGAGACAGATTATGGTTTACCTGGATGTGAGACACTTCTCCAGAGGAACATCAGATT 399

QY 38 LysValGlnAspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHis 57
Db 400 AAGTAGCAGATGACTATCTGCAGATTTCATGCAAGCAGTGGAGAGAACAGACGACCAT 459

QY 58 GlyTyrIleSerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSer 77
Db 460 GGCTACATCTCAGCGGAGTTCACCGCCGCTACCGCTCTACCTCTCAACGTCGATCAGGCA 519

QY 78 AlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGln 97
Db 520 GCCATCACTTGTACGCTGTCACTGATGGCTGTCTCAGCTTCTGGGCCCCAAA---ACA 576

QY 98 ThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysPro 117
Db 577 GCGCGATCAAAATACGGCGTGGAGATCGCACCAATCTCTGTCAACCGCGATGACAAACCC 636

QY 118 ThrSerAlaProSerSer 123
Db 637 AACCTGCTGCTCTCTCT 654

RESULT 28
OLJ000940 436 bp mRNA linear VRT 02-MAY-1998
LOCUS Oryzias latipes, alpha-A-crystallin gene.
ACCESSION AJ000940
VERSION AJ000940.1 GI:3115327
KEYWORDS alpha-A-crystallin gene.
SOURCE Oryzias latipes (Japanese medaka)
ORGANISM Oryzias latipes
REFERENCE 1
AUTHORS Loosli,F., Koster,R.W., Carl,M., Krone,A. and Wittbrodt,J.
TITLE Six3, a medaka homologue of the Drosophila homeobox gene sine
oculus is expressed in the anterior embryonic shield and the
developing eye
JOURNAL Mech. Dev.
VOLUME 2 (bases 1 to 436)
AUTHORS Loosli,F.
TITLE Direct Submission
JOURNAL Submitted (12-FEB-1998) Loosli F., Institute for Human Genetics,
c/o MPI of Biophys. Chem., University of Gottingen, Am Fassberg,
Gottingen, 37077, GERMANY
FEATURES
Location/Qualifiers
source
1. 436
/organism="Oryzias latipes"
/mol_type="mRNA"
/strain="Carolina Biological"
/db_xref="taxon:8090"
1. 436
/gene="alpha-A-crystallin"
<1..>436
/gene="alpha-A-crystallin"
/protein_id="CAA04397.1"
/codon_start=1
/db_xref="GI:3115328"
/db_xref="GOA:O73919"
/db_xref="UniProt/Swiss-Prot:O73919"
/translation="RLFDQFEGMFHDLLPFTSPITSPYRQSLFRNFDLSSNSGI
SEVSRDRDKFTVHWDVHKHFSDELISKVVIDDFVIOGKHGERQDDHGYISREFFHRYR

LPSTWDSQSAITCSLSADGLLTLLSPNPAGGNRSDRSIPVCR"

ORIGIN
Alignment Scores:
Pred. No.: 4.99e-37 Length: 436
Score: 408.00 Matches: 82
Percent Similarity: 80.17% Conservative: 11
Best Local Similarity: 70.69% Mismatches: 19
Query Match: 63.95% Indels: 4
DB: 5 Gaps: 2

US-10-657-740-1_COPY_51_173 (1-123) x OLJ000940 (1-436)
QY 1 SerLeuPheArgThrValLeuAsp-----SerGlyIleSerGluValArgSerAsp 17
Db 91 TCGTGTGTTCTTAATCTCTCGACTCTCTCAAACTCCGGCATTTCCGAGGTGAGGTCTGAC 150

QY 18 ArgAspLysPheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrVal 37
Db 151 AGGACAAAGTTTACCGTTCACCTGGATGTCAAGCACTTCTCCCCGATGAGCTCAGCGTG 210

QY 38 LysValGlnAspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHis 57
Db 211 AAGGTGATCGATGACTTTGTGGAGATCCAGGCGCAAGCATGCGAGAAAGACAGATGACCAT 270

QY 58 GlyTyrIleSerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSer 77
Db 271 GGCTACATCTCCCGGAGTTTACCGCCGCTACCGCTCTCTCTCTCTCTCTCTCTCTCTCT 330

QY 78 AlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGln 97
Db 331 GCCATCACTGCTCTCTCTCTCTCGATGGACTGTGACCTGTCCGGGCCAAACCCAGCG 390

QY 98 ThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProValSerArg 113
Db 391 GGGGGCGCGAATGGA---CGCAGCGACCGCAGCATCCCGTCTGCCGC 435

RESULT 29
LOCUS RATACRYA 249 bp mRNA linear ROD 27-APR-1993
DEFINITION Rat alpha A-crystallin mRNA, partial cds.
ACCESSION M96949
VERSION M96949.1 GI:202619
KEYWORDS alpha-A-crystallin.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
REFERENCE 1 (bases 1 to 249)
AUTHORS Srinivasan,A.N., Nagineni,C.N. and Bhat,S.P.
TITLE Alpha A-crystallin is expressed in non-ocular tissues
JOURNAL J. Biol. Chem. 267 (32), 23337-23341 (1992)
MEDLINE 93054670
PUBMED 1429679
COMMENT Original source text: Rattus norvegicus (strain Sprague-Dawley)
neonate spleen cDNA to mRNA.
FEATURES
Location/Qualifiers
source
1. 249
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/tissue_type="spleen"
/dev_stage="neonate"
1. 249
/gene="alpha A-crystallin"
<1..>249
/gene="alpha A-crystallin"
/notes="this sequence is 100% homologous to previously
reported sequence of rat eye alpha A-crystallin mRNA-cDNA"
/codon_start=1
/product="alpha A-crystallin"
gene
CDS

```

```

/protein_id="AAA40645.1"
/db_xref="GI:202622"
/translation="LEDVFEHGKHNERQDDHGYISREFFHYRLPSNMDOQALSCL
SADGMTPTSGPKVSGI.DACHSRRLPVSRPKPSAPS"

```

## ORIGIN

Alignment Scores:		
Pred. No.:	7,378-37	Length:
Score:	404.00	Matches:
Percent Similarity:	97.56%	Conservative:
Best Local Similarity:	91.46%	Mismatches:
Query Match:	63.22%	Indels:
DB:	10	Gaps:
		249
		75
		2
		0
		0

US-10-657-740-1 COPY 51 173 (1-123) X RATAACRYA (1-249)

Qy	41	AspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIle	60
		:::	
Db	4	GAAGATTTCGTGCAGATCCATGGCAAAACACAACAGAGGCGAGATGACCATGGCTACATT	63
Qy	61	SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer	80
		:::	
Db	64	TCCCGTGAATTTACCGTCGTACCGTCTGCTCTCCAATGFGACCAAGTCCGCCCTCTCC	123
Qy	81	CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu	100
		:::	
Db	124	TGTCCTCTGTCGCGATGGCATGCTGACCTTCTCTGSCCCCAAGGTCACAGTCTGGCTTG	183
Qy	101	AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGlyLysProThrSerAla	120
		:::	
Db	184	GATGCTGGCCACAGCGAGAGGCGCCATTCCTCGTGTACGGGAGGAGAACCCAGCTCGGCA	243
Qy	121	ProSer	122
Db	244	CCCTCG	249

RESULT 30	RATAACRYB	249 bp	linear	27-APR-1993
LOCUS	Rat alpha A-crystallin mRNA, partial cds.			
DEFINITION				
ACCESSION	M96950			
VERSION	M96950.1	GI:202621		
KEYWORDS	alpha-A-crystallin.			
SOURCE	Rattus norvegicus (Norway rat)			
ORGANISM	Rattus norvegicus			
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;			
	Rattus.			
REFERENCE	1 (bases 1 to 249)			
AUTHORS	Srinivasan,A.N., Nagineni,C.N. and Bhat,S.P.			
TITLE	alpha A-crystallin is expressed in non-ocular tissues			
JOURNAL	J. Biol. Chem. 267 (32), 23337-23341 (1992)			
MEDLINE	93054670			
PUBLISHED	1429679			
COMMENT	Original source text: Rattus norvegicus (strain Sprague-Dawley) neonate eye cDNA to mRNA.			

```

FEATURES
source
1. .249
Location/Qualifiers
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/tissue_type="eye"
/dev_stage="neonate"
1. .249
gene
/gene="alpha A-crystallin"
<1. .>249
CDS
/gene="alpha A-crystallin"
/note="this sequence shows a single base change (bp 103)
with respect to M96349 which alters amino acid from valine
to methionine"
/codon_start=1
/product="alpha A-crystallin"

```

## ORIGIN

	Pred. No.: Score: Percent Similarity: Best Local Similarity: Query Match: DB:	1.61e-36 401.00 97.5% 90.24% 62.75% 10	Length: Matches: Conservative: Mismatches: Indels: Gaps:	249 74  2 0 0
--	--	---	---	------------------------------

US-10-657-740-1 COPY 51 173 (1-123) X RATAACRYB (1-249)

Qy	41	AspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAepHisGlyTyrIle	60
		:::	
Db	4	GAAGATTTCGTGAGATCCATGCGAAACAAACGAGAGCGAGTACCATTCATT	63
Qy	61	SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer	80
		:::	
Db	64	TCCCGTGAAATTCACCGTCGCTACCGTCTCCATTGCAATATGACACAGTCCGCCCTCTCC	123
Qy	81	CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu	100
		:::	
Db	124	TGTCCTCTTGCTCGGGATGGCATGCTGACCTTCTCTGGCCCCAAGGTCCAGTCTGGCTTG	183
Qy	101	AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGlnLysProThrSerAla	120
		:::	
Db	184	GATGCTGGCCACAGCGAGGGGCCATTCCCGTGTACGGGAGGAGAACCCAGCTCGGCA	243
Qy	121	ProSer	122
Db	244	CCCTCG	249

RESULT 31	AF026952	419 bp	DNA	linear	PRI'02-APR-2001
LOCUS	AF026952				
DEFINITION	Homo sapiens alpha-A-crystallin (CRYAA) gene, partial cds.				
ACCESSION	AF026952				
VERSION	AF026952.1	GI:2827908			
KEYWORDS	.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	Litt,M., Kramer,P., LaMorticella,D.M., Murphey,W., Lovrien,E.W. and				
AUTHORS	Weleber,R.G.				
	Autosomal dominant congenital cataract associated with a missense				
TITLE	mutation in the human alpha crystallin gene CRYAA				
JOURNAL	Hum. Mol. Genet. 7 (3), 471-474 (1998)				
MEDLINE	98133928				
PubMed	9467006				
REFERENCE	2 (bases 1 to 419)				

```

AUTHORS      Litt,M.
TITLE        Direct Submission
JOURNAL      Submitted (25-SEP-1997) Molecular and Medical Genetics, Oregon
              Health Sciences University, 3181 S.W. Sam Jackson Park Rd.,
              Portland, OR 97201-3098, USA
FEATURES     Location/Qualifiers
              1..419
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /chromosome="21"
                /map="21q22.3"
              <1..>419
                /gene="CRYAA"
              <1..119
                /gene="CRYAA"
                /number=2
              gene
              intron

```

```

mRNA
<120. .>419
/gene="CRYAA"
/product="alpha-A-crystallin"
120. .>419
/gene="CRYAA"
/number=3
<120. .>329
/gene="CRYAA"
/codon_start=1
/product="alpha-A-crystallin"
/protein_id="AA99912.1"
/db_xref="GI:2827909"
/translators="DDHGYISREFHRRYRLPSNVDSALSCSLSDGMLTFCGPKIQT
GLDATAERAIPVSRREKPTSPSS"
330. .>419
/gene="CRYAA"

3'UTR

ORIGIN
Alignment Scores:
Pred. No.: 2,14e-32 Length: 419
Score: 367.00 Matches: 70
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 57.43% Indels: 0
DB: 9 Gaps: 0

US-10-657-740-1_COPY_51_173 (1-123) x AF026952 (1-419)

QY 54 GlnAspAspHisGlyTyrIleSerArgGluPheHisArgTyrArgLeuProSerAan 73
Db 117 CAGGACGACACCGCTACATTTCCCGTGAGTTCACCGCCGCTACCGCTGCCGCTCCAAC 176
QY 74 ValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGly 93
Db 177 GTGGACCAAGTCGGCCCTCTTGTCTCCCTGCTGCGATGGCATGCTGACCTTCTGTGGC 236
QY 94 ProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProValSerArg 113
Db 237 CCCAAGATCCAGATCGGCTGAGTGCACCCAGCGAGCGACCATCCCGCTGTGCGGG 296
QY 114 GluGluLysProThrSerAlaProSerSer 123
Db 297 GAGGAGAAGCCACCTCGGCTCCCTCGTCC 326

RESULT 32
AP001631
LOCUS
DEFINITION
Homo sapiens genomic DNA, chromosome 21, clone:KB2007G4,
AP001631
ACCESSION
AP001631.1 GI:7670585
VERSION
2 (bases 1 to 148179)
KEYWORDS
HTG.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 148179)
AUTHORS
Shimizu,N., Kudoh,J. and Shibuya,K.
TITLE
Homo sapiens genomic DNA, chromosome 21, clone:KB2007G4,
JOURNAL
Published Only in DataBase (2000)
REFERENCE
2 (bases 1 to 148179)
AUTHORS
Shimizu,N., Kudoh,J. and Shibuya,K.
TITLE
Direct Submission
JOURNAL
Submitted (04-APR-2000) Nobuyoshi Shimizu, Keio University, School
of Medicine, Molecular Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo
160-8582, Japan (E-mail:nshimizu@db.med.keio.ac.jp,
Tel:81-3-3351-2370, Fax:81-3-3351-2370)
FEATURES
Location/Qualifiers
1..148179
source
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

```

```

/chromosome="21"
/clone="KB2007G4"
/cell_line="FLEB14-14"
/cell_type="pre-pro-B cell"
/clone_lib="Keio BAC library"

```

## ORIGIN

```

Alignment Scores:
Pred. No.: 2.05e-29 Length: 148179
Score: 367.00 Matches: 70
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 57.43% Indels: 0
DB: 9 Gaps: 0

US-10-657-740-1_COPY_51_173 (1-123) x AP001631 (1-148179)

QY 54 GlnAspAspHisGlyTyrIleSerArgGluPheHisArgTyrArgLeuProSerAan 73
Db 132213 CAGGACGACACCGCTACATTTCCCGTGAGTTCACCGCCGCTACCGCTGCCGCTCCAAC 132272
QY 74 ValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGly 93
Db 132273 GTGGACCAAGTCGGCCCTCTTGTCTCCCTGCTGCGATGGCATGCTGACCTTCTGTGGC 132332
QY 94 ProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProValSerArg 113
Db 132333 CCCAAGATCCAGATCGGCTGAGTGCACCCAGCGAGCGACCATCCCGCTGTGCGGG 132392
QY 114 GluGluLysProThrSerAlaProSerSer 123
Db 132393 GAGGAGAAGCCACCTCGGCTCCCTCGTCC 132422

RESULT 33
BS000233
LOCUS
DEFINITION
Pan troglodytes chromosome 22 clone:CH251-445J22, map 22, complete
sequences.
ACCESSION
BS000233 BA000046
VERSION
BS000233.1 GI:37537500
KEYWORDS
HTG.
SOURCE
Pan troglodytes (chimpanzee)
ORGANISM
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE
1
AUTHORS
The International Chimpanzee Chromosome 22 Consortium.
TITLE
DNA sequence and comparative analysis of chimpanzee chromosome 22
JOURNAL
Nature 429, 382-388 (2004)
REFERENCE
2 (bases 1 to 163962)
AUTHORS
Tsai,S., Liu,T., Wu,K., Liao,T. and Hsiao,K.
TITLE
Direct Submission
JOURNAL
Submitted (16-MAY-2003) Shih-Feng Tsai, National Health Research
Institutes (NHRI), Division of Molecular and Genomic Medicine; 128,
Yen-Chiu-Yuan Road, Sec 2, Taipei 115, Taiwan
(E-mail:petsei@nhri.org.tw, URL:http://www.nhri.org.tw/,
Tel:886-2-28267319, Fax:886-2-28200552)
COMMENT
*Chinese National Human Genome Center at Shanghai, Shanghai, China;
*CBF, Dept. of Genome Analysis, Braunschweig, Germany; *Institute
of Molecular Biotechnology, Jena, Germany; *KRIIBB Genome Research
Center, Daejeon, Korea;
*Max-Planck-Institute for Molecular Genetics, Berlin, Germany;
*National Institute of Genetics, Mishima, Japan;
*National Yang Ming University Genome Research Center, Taipei,
Taiwan;
*RIKEN Genomic Sciences Center, Yokohama, Japan.
----- Genome Center
Center: National Yang Ming University Genome Research Center
code: YMGCC
Web site: http://genome.ym.edu.tw/
Contact: sequence@ym.edu.tw
----- Project Information

```



Center project name: The Chimpanzee Chromosome 22 Sequencing Project  
Center clone name: HX

#### ----- Summary Statistics

Sequencing vector: pUC18; 100% of reads  
Chemistry: Dye-terminator Big Dye and ET; 100% of reads Assembly  
Program: Phrap; version 0.990319  
Consensus quality: 162,928 bases at least Q40  
Consensus quality: 163,029 bases at least Q30  
Consensus quality: 163,039 bases at least Q20

-----  
This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30);  
an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at one plasmid  
subclone or more than one M13 subclone;  
and the assembly was confirmed by restriction digest.

#### ----- Source information:

The CHORI-251 chimpanzee BAC library was prepared from DNA isolated from the blood of a single male chimpanzee using published protocols (Osoegawa, K. et al. Genomics 52:1-8). The DNA from the chimpanzee ('Clint') was obtained from the Yerkes Primate Center in Atlanta. The library was prepared by Baoli Zhu, Chung Li Shu, Kazutoyo Osoegawa, Evan Eichler & Pieter J de Jong. The library characteristics are described at  
<http://www.chori.org/bacpac/chimpanzee251.htm>.  
The clone may be obtained from Pieter J. de Jong and coworkers  
(<http://www.chori.org/bacpac>).  
VECTOR: pIARBAC2.1

#### ----- Sequence Quality Assessment:

This entry has been annotated with sequence estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than 1 error in 10,000 bp.

-----  
Neighboring clones: PTB-071B01(left) and PTB-190I13(right).

#### FEATURES

source  
1..163962  
Location/Qualifiers  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
/chromosome="22"  
/clone\_lib="CH251-445J22"  
/note="low quality region"  
76772..76776  
/note="gap containing unresolved di-nucleotide repeats, (CA)n"  
misc\_feature  
138531..138532  
/note="low quality region"  
misc\_feature  
138537  
/note="low quality region"  
misc\_feature  
138539..138540  
/note="low quality region"

#### ORIGIN

Alignment Scores:  
Pred. No.: 2,31e-29 Length: 163962  
Score: 367.00 Matches: 70  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 57.43% Indels: 0  
DB: 9 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x BS000233 (1-163962)

Qy 54 GlnAspAspHisGlyTyrIleSerArgGluPheHisArgTyrArgLeuProSerAsn 73

Db 135934 CAGGATGACCGGCTACATTTCCCGTGAGTTCCACCGCGCTACCGCTGCCGTCACAC 135993  
Qy 74 ValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGly 93  
Db 135994 GTGGACCACTGGCCCTCTCTTGCTCCCTGTCCGCGATGGCATGCTGACCTTCTGTGGC 136053  
Qy 94 ProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProValSerArg 113  
Db 136054 CCCAGATCCAGACTGGCTGGATGCCACCCACGCGAGCGAGCCATCCCGCTGTGCGCG 136113  
Qy 114 GluGluLysProThrSerAlaProSerSer 123  
Db 136114 GAGGAGAGGCCACCCTCGCTCCCTCGTCC 136143

#### RESULT 34

AP001748  
LOCUS AP001748 340000 bp DNA linear PRI 21-MAY-2003  
DEFINITION Homo sapiens genomic DNA, chromosome 21q, section 92/105.  
ACCESSION AP001748 AL163293 BA000005  
VERSION AP001748.1 GI:7768745  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1  
AUTHORS Hattori, M., Fujiyama, A., Taylor, T.D., Watanabe, H., Yada, T., Park, H.S., Toyoda, A., Ishii, K., Totoki, Y., Choi, D.K., Soeda, E., Ohki, M., Takagi, T., Sakaki, Y., Taudien, S., Blechschmidt, K., Polley, A., Menzel, U., Delabar, J., Kumpf, K., Lehmann, R., Patterson, D., Reichwald, K., Rump, A., Schillhabel, M., Schudy, A., Zimmermann, W., Rosenthal, A., Sasaki, K., Nagamine, K., Mitsuyma, S., Asakawa, S., Shintani, A., Sasaki, K., Kudoh, J., Shibuya, K., Kawasaki, A., Antonarakis, S.E., Minoshima, S., Shimizu, N., Nordstreck, G., Hornischer, K., Barandt, P., Scharfe, M., Schoen, O., Desario, A., Reichelt, J., Kauer, G., Bloeker, H., Ramser, J., Beck, A., Klages, S., Hennig, S., Resselmann, L., Dagand, E., Wehrmeyer, S., Borzym, K., Gardiner, K., Nizetic, D., Francis, F., Lehrach, H., Reinhardt, R. and Yaspo, M.L.  
The DNA sequence of human chromosome 21

#### TITLE

JOURNAL Nature 405 (6784), 311-319 (2000)  
MEDLINE 20289799  
PUBMED 10830953

#### REFERENCE

2 (bases 1 to 340000)

#### AUTHORS

Hattori, M., Fujiyama, A., Taylor, T.D., Watanabe, H., Yada, T., Park, H.S., Toyoda, A., Ishii, K., Totoki, Y., Choi, D.K., Soeda, E., Ohki, M., Takagi, T., Sakaki, Y., Taudien, S., Blechschmidt, K., Polley, A., Menzel, U., Delabar, J., Kumpf, K., Lehmann, R., Patterson, D., Reichwald, K., Rump, A., Schillhabel, M., Schudy, A., Zimmermann, W., Rosenthal, A., Sasaki, K., Kudoh, J., Shibuya, K., Kawasaki, A., Antonarakis, S.E., Minoshima, S., Shimizu, N., Nordstreck, G., Hornischer, K., Barandt, P., Scharfe, M., Schoen, O., Desario, A., Reichelt, J., Kauer, G., Bloeker, H., Ramser, J., Beck, A., Klages, S., Hennig, S., Resselmann, L., Dagand, E., Wehrmeyer, S., Borzym, K., Gardiner, K., Nizetic, D., Francis, F., Lehrach, H., Reinhardt, R. and Yaspo, M.L.

#### TITLE

JOURNAL

Submitted (10-APR-2000) The Chromosome 21 Mapping and Sequencing Consortium: \* RIKEN Genomic Sciences Center, Human Genome Research Group \* Institute of Molecular Biotechnology, Genome Analysis \* Keio University School of Medicine, Dept. of Molecular Biology \* GBF, Dept. of Genome Analysis \* Max-Planck Institute for Molecular Genetics (addresses see below)

#### COMMENT

On May 30, 2000 this sequence version replaced gi:7717405.  
The chromosome 21 mapping and sequencing consortium consisting of  
\* RIKEN Genomic Sciences Center, Human Genome Research Group, \* Sagami-hara 228-8555, Japan.  
\* e-mail: hattori@gsc.riken.go.jp  
\* URL: <http://hgp.gsc.riken.go.jp/>  
and  
\* Institute of Molecular Biotechnology, Genome Analysis, \*



Pred. No.: 5,428-29 Length: 340000  
 Score: 367.00 Matches: 70  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 57.43% Indels: 0  
 DB: 9 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x AP001748 (1-340000)

Qy 54 GlnAspAspHisGlyTyrIleSerArgGluPheHisArgTyrArgLeuProSerAsn 73  
 Db 327926 CAGGACGACACCGGTACATTCCCGTGAGTTCCACACCGCGCTACCGCTGCGTCCCAAC 327985  
 Qy 74 ValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGly 93  
 Db 327986 GTGGACCAAGTCGGCCCTCTTGCTCCCTGCTGTCGCGATGCGATGCTGACCTTCTGTGGC 328045  
 Qy 94 ProIlyIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProValSerArg 113  
 Db 328046 CCCAAGATCCAGATCGGCTGATGCCACCCAGCGAGCGACCATCCCCGTGTGCGG 328105  
 Qy 114 GluGluLysProThrSerAlaProSerSer 123  
 Db 328106 GAGGAGAGCCACCTCGGCTCCCTCGTCC 328135

## RESULT 35

AJ617732 AJ617732 507 bp mRNA linear MAM 15-JUL-2004  
 LOCUS Elephas maximus partial mRNA for alphaB-crystallin (cryab gene).  
 DEFINITION

ACCESSION AJ617732

VERSION AJ617732.1 GI:50344360

KEYWORDS alphaB-crystallin; cryab gene.

SOURCE Elephas maximus (Asiatic elephant)

## ORGANISM

Elephas maximus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Proboscidea; Elephantidae; Elephas.

## REFERENCE

1 van Rheede, T. and de Jong, W.W.

THE alpha-crystallins of the platypus Ornithorhynchus anatinus.  
 Origin of the alternatively spliced exon alphaAins and implications  
 for mammalian phylogeny

## JOURNAL

Unpublished

2 (bases 1 to 507)

## AUTHORS

Franc, B.

Direct Submission

## TITLE

Submitted (16-DEC-2003) Franck E., Biochemistry, Radboud University

## JOURNAL

Nijmegen, 9101, 6500 HB Nijmegen, NETHERLANDS

## FEATURES

Location/Qualifiers

1..507

/organism="Elephas maximus"

/mol\_type="mRNA"

/db\_xref="taxon:9783"

## gene

1..507

/gene="cryab"

<1..507

/gene="cryab"

/product="alphaB-crystallin"

/protein\_id="CAP02108.1"

/db\_xref="GI:50344361"

/translation="PWIRRRPFFPHSPSRLFDQFGEHLLSDLPFATSLSPFYLRP  
 PSFLRAPSFWFTGLSEMRLEKDRFSVNLVDKHFSPEDLKVGLGVDIEVHGHEERQD  
 EHGFIISREPHKRYIPADVDPLAITSSLSGDLTVNGPRKQAPGPRTIPIITREKQ  
 AVTAAPKK"

## ORIGIN

## Alignment Scores:

Pred. No.: 5,86e-32 Length: 507  
 Score: 364.00 Matches: 69  
 Percent Similarity: 79.31% Conservative: 23  
 Best Local Similarity: 59.48% Mismatches: 18  
 Query Match: 56.96% Indels: 6  
 DB: 4 Gaps: 2

US-10-657-740-1\_COPY\_51\_173 (1-123) x AJ617732 (1-507)  
 Qy 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27  
 Db 163 GACACTGGCTCTCAGAGATCGGCTGGAGAAGACAGAGTTCTCTGTCAACCTGGATGTG 222  
 Qy 28 LysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGluLeuHis 47  
 Db 223 AAGCATTTCTCTCCAGAGGATCTCAAGGTCAAGGTGCTGGAGATGTGTGAGGTGCAT 282  
 Qy 48 GlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArgArg 67  
 Db 283 GGCACCAATGAGAACGCCAGGATGAACATGTTTCATCTCCGGGAGTTTCCATAGGAAA 342  
 Qy 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87  
 Db 343 TACCGATCCAGCTGATGTGACCCCTCTTCCCATTTACTTCATCCCTGTCATCTGATGG 402  
 Qy 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107  
 Db 403 GTCTCTTACTGTGAATGGACCAAGGAAACAG-----GCCCTGGCCCTGAGCGC 450  
 Qy 108 AlaIleProValSerArgGluLysPro-----ThrSerAlaPro 121  
 Db 451 ACCATTCCCATCACCCGGTGAAGAGAGCCTGCTGTCATCGCAGCTCCC 498

## RESULT 36

AF029793

LOCUS Bos taurus alpha B-crystallin (CRYAB) mRNA, complete cds.

DEFINITION AF029793

VERSION AF029793.2 GI:5296002

KEYWORDS

SOURCE Bos taurus (cow)

ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 Bovinae; Bos.

REFERENCE 1 (bases 1 to 632)

AUTHORS Kelley, P.B., Abraham, E.C., Zhao, H.R., Shroff, N.P., Cherian, M. and Thomas, J.J.

Direct Submission

Submitted (10-OCT-1997) Biochemistry and Molecular Biology, Medical

College of Georgia, 1120 15th Street, Augusta, GA 30912, USA

REFERENCE 2 (bases 1 to 632)

AUTHORS Kelley, P.B., Abraham, E.C., Zhao, H.R., Shroff, N.P., Cherian, M. and Thomas, J.J.

Direct Submission

Submitted (30-JUN-1999) Biochemistry and Molecular Biology, Medical

College of Georgia, 1120 15th Street, Augusta, GA 30912, USA

REMARK Sequence update by submitter

COMMENT On Jun 30, 1999 this sequence version replaced gi:2760900.

FEATURES Location/Qualifiers

1..632

/organism="Bos taurus"

/mol\_type="mRNA"

/db\_xref="taxon:9913"

1..632

/gene="CRYAB"

1..528

/gene="CRYAB"

/product="alpha B-crystallin"

/protein\_id="AAB95323.2"

/db\_xref="GI:5296003"

/translation="MDIAIHPWIRRRPFFPHSPSRLFDQFGEHLLSDLPFATSLSPFYLRP  
 PSFLRAPSFWFTGLSEMRLEKDRFSVNLVDKHFSPEDLKVGLGVDIEVHGHEERQD  
 KHEERQDHEGFIISREPHKRYIPADVDPLAITSSLSGDLTVNGPRKQASGPRTIP  
 ITREKQAVTAAPKK"

gene

CDS

ORIGIN

Alignment Scores:

```

Pred. No.: 7.58e-32 Length: 632
Score: 364.00 Matches: 68
Percent Similarity: 80.34% Conservative: 26
Best Local Similarity: 58.12% Mismatches: 17
Query Match: 56.96% Indels: 6
DB: 4 Gaps: 2

US-10-657-740-1_COPY_51_173 (1-123) x AF029793 (1-632)

QY 7 LeuaspSerGlyIleSerGluValArgSerAspArgAspIysPheValIlePheLeuAsp 26
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 181 ATTGACACTGGCCCTCTCAGATCGCTGTGAGAAAGACAGATCTCTGTCAACCTGGAT 240
QY 27 VallysHisPheSerProGluAspLeuThrVallysValGlnAspAspPheValGluIle 46
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 241 GTGAAGCACTTCTCCAGAGAACTCAAGTCAAGTGTGGAGATGTGATGAGGTG 300
QY 47 HisGlyIysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArg 66
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 301 CATGGCAACATGAAGAGAGCCAGGATGAACATGGTTTATCTCCGGGAGTTCCACAGG 360
QY 67 ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 361 AAATACCGGATCCAGCTGACGTGGACCTCTCGCCATTACTTCCCTGCTCTGAT 420
QY 87 GlyMetLeuThrPheCysGlyProLyIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 421 GGGGTCTCTACTGTGAATGACCAAGAAACAG-----GCCTCCGGCCCTGAG 468
QY 107 ArgAlaIleProValSerArgGluGluIysPro-----ThrSerAlaPro 121
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 469 CGCACCATTCCTCCATCACCCTGTAAGAGAAAGCGCGGTGTCTACTGCAGCCCC 519

RESULT 37
S74229
LOCUS
DEFINITION
alpha B-crystallin [rats, heart, mRNA, 671 nt].
ACCESSION
S74229
VERSION
S74229.1 GI:241448
KEYWORDS
Rattus sp.
SOURCE
Rattus sp.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 671)
Atomi, Y., Yamada, S. and Nishida, T.
Early changes of alpha B-crystallin mRNA in rat skeletal muscle to
mechanical tension and denervation
JOURNAL
Biochem. Biophys. Res. Commun. 181 (3), 1323-1330 (1991)
MEDLINE
92109745
PubMed
1764082
REMARK
GenBank staff at the National Library of Medicine created this
entry [NCBI gi241448] from the original journal article.
FEATURES
Location/Qualifiers
1..671
/organism="Rattus sp."
/mol_type="mRNA"
/db_xref="taxon:10118"
1..671
/gene="alpha B-crystallin"
7..534
/gene="alpha B-crystallin"
/codon_start=1
/product="alpha B-crystallin"
/protein_id="AAB20759.1"
/db_xref="GI:241449"
translation="MDIAIHPTWRPFPPFPSPRLFDQFGHEHLESDLPSTATSL
SPFYRPFSLRAPSWIDTGLSEMRMEKDRFSVNLVDKHSPEELKVLGVIEVHG
KHREORDEHGFISREPRKRYIPADVDPDLTITSSLSDSGVLTVNGPRKQASGPFTIP
ITREKPAVTAAPKK"
gene
CDS
ORIGIN

```

```

Alignment Scores:
Pred. No.: 1.78e-31 Length: 671
Score: 361.00 Matches: 67
Percent Similarity: 79.43% Conservative: 26
Best Local Similarity: 57.26% Mismatches: 18
Query Match: 56.49% Indels: 6
DB: 10 Gaps: 2

US-10-657-740-1_COPY_51_173 (1-123) x S74229 (1-671)

QY 7 LeuaspSerGlyIleSerGluValArgSerAspArgAspIysPheValIlePheLeuAsp 26
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 187 ATTGACACTGGCCCTCTCAGATCGCTGTGAGAAAGACAGATCTCTGTGAACCTGGAC 246
QY 27 VallysHisPheSerProGluAspLeuThrVallysValGlnAspAspPheValGluIle 46
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 247 GTGAAGCACTTCTCTCAGAGAACTCAAGTCAAGTGTCTGGAGAGCGTGAATTGAGGTG 306
QY 47 HisGlyIysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArg 66
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 307 CACGGCAAGCAGCAAGAGCGCCAGGACCAATCGCTTCTATCTCCAGGAGTTCACAGG 366
QY 67 ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 367 AAGTACCGGATCCAGCGGAGCTGATCCTCTCACCATTACTTCTTCCCTGTCTCGGAT 426
QY 87 GlyMetLeuThrPheCysGlyProLyIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 427 GGAGTCTCTACTGTGAATGACCAAGAAACAG-----GCCTTCGGCCCTGAG 474
QY 107 ArgAlaIleProValSerArgGluGluIysPro-----ThrSerAlaPro 121
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 475 CGCACCATTCCTCCATCACCCTGTAAGAGAAAGCGGTGTCTACTGCAGCCCCCT 525

RESULT 38
RRHARTABC
LOCUS
DEFINITION
R. rattus mRNA for alpha B-crystallin (adult heart).
ACCESSION
X60352
VERSION
X60352.1 GI:57579
KEYWORDS
alpha B-crystallin.
Rattus rattus (black rat)
SOURCE
Rattus rattus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 687)
Bhat, S.P., Horwitz, J., Srinivasan, A. and Ding, L.
alpha B-crystallin exists as an independent protein in the heart and
in the lens
JOURNAL
Eur. J. Biochem. 102, 775-781 (1991)
REFERENCE
2 (bases 1 to 687)
Bhat, S.P.
Direct Submission
Submitted (25-SEP-1991) S.P. Bhat, Room B-118, Jules Stein Eye
Institute, UCLA School of Medicine, 100 Stein Plaza, Los Angeles,
California 90024-7008, USA
A multimeric ubiquitous extracellular crystallin of the vertebrate
lens, related to small heat shock proteins, phosphorylated.
Presumed function: structural protein of the lens in association
with alpha A-crystallin.
FEATURES
Location/Qualifiers
1..687
/organism="Rattus rattus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10117"
/clone="p alpha RH8"
/cell_line="adult heart"
/tissue type="heart"
/clone_id="lambda qt11"
/dev_stage="adult"
<1..525

```

CDS

```

/codon_start=1
/product="alpha B-crystallin"
/protein_id="CAA42911.1"
/db_xref="GI:57580"
/db_xref="GOA:P23928"
/db_xref="UniProt/Swiss-Prot:P23928"
/translation="MDIAIHPWIRPPFPSPRLDPQFFGEHLLSDLFSTATSL
PFYLRPSPFLAPSWIDTGLSEMRMEKDRFSVNLVDVKHFSPEELKVGLGVIEVHGK
HEERQDEHGFISREFHRYKIPADVDPLTITSSLSGDLTVNGPRKQASGPERTIP
TREKPAVTAAPKK"
polyA_signal
625..630
ORIGIN
Alignment Scores:
Pred. No.: 1.83e-31 Length: 687
Score: 361.00 Matches: 67
Percent Similarity: 79.49% Conservative: 26
Best Local Similarity: 57.26% Mismatches: 18
Query Match: 56.49% Indels: 6
DB: 10 Gaps: 2
US-10-657-740-1_COPY_51_173 (1-123) x RHARTABC (1-687)
Qy 7 LeuAspSerGlyIleSerGluValArgSerArgAspLysPheValIlePheLeuAsp 26
Db 178 ATTGACCTGGCTCTCAGATGCGTATGAGAGGACAGGTTCTCTGTGAACCTGGAC 237
Qy 27 VallysHisPheSerProGluAspLeuThrVallysValGlnAspPheValGluIle 46
Db 238 GTGAAGCACTTCTCTCCAGAGAACTCAAGTCAAGTTCTGGGAGAGCGTATTGAGGTG 297
Qy 47 HisGlyLysHisAsnGluArgGlnAspHisGlyTyrlleSerArgGluPheHisArg 66
Db 298 CACGCAAGCAGCAGAGCGGACGAGCAATGCTTCTCATCTCCAGGAGTTCACAGG 357
Qy 67 ArgTyArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86
Db 358 AAGTACCGGATCCAGCGGACGATGATCTCTCACCATTACTTCTTCCCTGTCTCGGAT 417
Qy 87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106
Db 418 GGAGTCTCTGCTGCAATGGACCAAGAAACAG-----GCCTCTGGCCCTGAG 465
Qy 107 ArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 121
Db 466 CGCACCATTCACATCACCCTGGAAGAGAGCGCTGTGTCTGCTGACGAGCCCT 516
RESULT 39
RLENSABC 689 bp mRNA linear ROD 31-JAN-1992
LOCUS R.rattus mRNA for alpha B-crystallin (ocular lens tissue).
DEFINITION X60351
ACCESSION X60351
VERSION 1
KEYWORDS alpha-B-crystallin.
SOURCE Rattus rattus
ORGANISM Rattus rattus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 689)
Bhat,S.P., Horwitz,J., Srinivasan,A. and Ding,L.
alphaB-crystallin exists as an independent protein in the heart and
in the lens
Eur. J. Biochem. 102, 775-781 (1991)
2 (bases 1 to 689)
Bhat,S.P.
Direct Submission
Submitted (25-SEP-1991) S.P. Bhat, Room B-118, Jules Stein Eye
Institute, UCLA School of Medicine, 100 Stein Plaza, Los Angeles,
California 90024-7008, USA
A multimeric ubiquitous extracellular crystallin of the vertebrate
lens, related to small heat shock proteins, phosphorylated.
Presumed function: structural protein of the lens in accommodation
REFERENCE 1 (bases 1 to 704)
RLENSABC 704 bp mRNA linear ROD 06-MAY-2003
LOCUS S77142
DEFINITION alpha B-crystallin [rats, heart, mRNA, 704 nt].
ACCESSION S77142
VERSION S77142.1 GI:242298
KEYWORDS Rattus sp.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 704)

```

**AUTHORS** Bhat,S.P., Horwitz,J., Srinivasan,A. and Ding,L.  
**TITLE** Alpha B-crystallin exists as an independent protein in the heart  
**JOURNAL** Eur. J. Biochem. 202 (3), 775-781 (1991)  
**MEDLINE** 92111524  
**PUBMED** 1765091  
**REMARK** GenBank staff at the National Library of Medicine created this entry [NCBI gibbsq 77142] from the original journal article.

**FEATURES** source  
 1..704  
 /organism="Rattus sp."  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10118"  
 <1..533  
 /codon\_start=3  
 /product="alpha B-crystallin"  
 /protein\_id="AAP31996.1"  
 /db\_xref="GI:30387800"  
 /translation="IPDIAIHPWIRPPFPSPRLFDQFGEHLLSDFSTATSL  
 LSPFYLRLPSPFLRAPSWIDTGLSEMRMEKDRFSVNLVDVKHFSPELKVGLGDIIEVHG  
 GKHEERQDEHGFISREHFKYRI PADVDPLTITSSLSDDGVLTVNGPRKQASGPRTIP  
 PITREKPAVTAAPKK"

## CDS

Alignment Scores:  
 Pred. No.: 1.88e-31 Length: 704  
 Score: 361.00 Matches: 67  
 Percent Similarity: 79.49% Conservative: 26  
 Best Local Similarity: 57.26% Mismatches: 18  
 Query Match: 56.49% Indels: 6  
 DB: 10 Gaps: 2

## ORIGIN

US-10-657-740-1\_COPY\_51\_173 (1-123) x S77142 (1-704)  
 QY 7 LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26  
 Db 186 ATTGACATCGGGCTCTCAGAGATCGGTATGGAGAGGACAGGTTCTCTGTGAACCTGGAC 245  
 QY 27 ValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGluIle 46  
 Db 246 GTGAGGACCTTCTCTCCAGAGAACTCAAGTCAAGTCTCTGGAGAGCGTATGAGGTG 305  
 QY 47 HisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArg 66  
 Db 306 CACGGCAAGCAGCAGAGCGCCAGGACCAATGCTTCATCTCCAGGAGTTCACACAGG 365  
 QY 67 ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86  
 Db 366 AAGTACCGGATCCAGCGGACGTGATCTCTCACCATTACTTCTTCCTGTCTATCGGAT 425  
 QY 87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106  
 Db 426 GGAGTCTCTCACTGTGAATGACCAAGAAACAG-----GCCCTCTGGCCCTGAG 473  
 QY 107 ArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 121  
 Db 474 GCACCATTCCTCATCCCGGTGAAGAGAGCGCTGTCTCACTGCAGGCCCT 524

**RESULT 41**  
**S77138**  
**LOCUS** alpha B-crystallin [rats, lens, mRNA, 706 nt]. linear ROD 06-MAY-2003  
**ACCESSION** S77138  
**VERSION** S77138.1 GI:242297  
**KEYWORDS**  
**SOURCE** Rattus sp.  
**ORGANISM** Rattus sp.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

**REFERENCE** 1 (bases 1 to 706)  
**AUTHORS** Bhat,S.P., Horwitz,J., Srinivasan,A. and Ding,L.  
**TITLE** Alpha B-crystallin exists as an independent protein in the heart

**JOURNAL** Eur. J. Biochem. 202 (3), 775-781 (1991)  
**MEDLINE** 92111524  
**PUBMED** 1765091  
**REMARK** GenBank staff at the National Library of Medicine created this entry [NCBI gibbsq 77138] from the original journal article.

**FEATURES** source  
 1..706  
 /organism="Rattus sp."  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10118"  
 21..548  
 /codon\_start=1  
 /product="alpha B-crystallin"  
 /protein\_id="AAP31995.1"  
 /db\_xref="GI:30387799"  
 /translation="MDIAIHPWIRPPFPSPRLFDQFGEHLLSDFSTATSL  
 SPFYLRLPSPFLRAPSWIDTGLSEMRMEKDRFSVNLVDVKHFSPELKVGLGDIIEVHG  
 KHEERQDEHGFISREHFKYRI PADVDPLTITSSLSDDGVLTVNGPRKQASGPRTIP  
 ITREKPAVTAAPKK"

## CDS

Alignment Scores:  
 Pred. No.: 1.89e-31 Length: 706  
 Score: 361.00 Matches: 67  
 Percent Similarity: 79.49% Conservative: 26  
 Best Local Similarity: 57.26% Mismatches: 18  
 Query Match: 56.49% Indels: 6  
 DB: 10 Gaps: 2

## ORIGIN

US-10-657-740-1\_COPY\_51\_173 (1-123) x S77138 (1-706)  
 QY 7 LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26  
 Db 201 ATTGACATCGGGCTCTCAGAGATCGGTATGGAGAGGACAGGTTCTCTGTGAACCTGGAC 260  
 QY 27 ValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGluIle 46  
 Db 261 GTGAGGACCTTCTCTCCAGAGAACTCAAGTCAAGTCTCTGGAGAGCGTATGAGGTG 320  
 QY 47 HisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArg 66  
 Db 321 CACGGCAAGCAGCAGAGCGCCAGGACCAATGCTTCATCTCCAGGAGTTCACACAGG 380  
 QY 67 ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86  
 Db 381 AAGTACCGGATCCAGCGGACGTGATCTCTCACCATTACTTCTTCCTGTCTATCGGAT 440  
 QY 87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106  
 Db 441 GGAGTCTCTCACTGTGAATGACCAAGAAACAG-----GCCCTCTGGCCCTGAG 488  
 QY 107 ArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 121  
 Db 489 CGCACCATTCCTCATCCCGGTGAAGAGAGCGCTGTCTCACTGCAGGCCCT 539

**RESULT 42**  
**AX401743**  
**LOCUS** Sequence 1419 from Patent WO0210453. linear PAT 06-JUN-2002  
**DEFINITION** AX401743  
**ACCESSION** AX401743  
**VERSION** AX401743.1 GI:21337923  
**KEYWORDS**  
**SOURCE** Rattus norvegicus (Norway rat)  
**ORGANISM** Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

**REFERENCE** 1  
**AUTHORS** Mendrick,D., Porter,M.W., Johnson,K.R., Castle,A.L. and Elashoff,M.R.  
**TITLE** Molecular toxicology modeling  
**JOURNAL** Patent: WO 0210453-A 1419 07-FEB-2002;

```
FEATURES
  source
    Gene Logic, Inc. (US)
      Location/Qualifiers
        1..1247
          /organism="Rattus norvegicus"
          /mol_type="unassigned DNA"
          /db_xref="taxon:10116"
          /note="EMBL/GenBank Accession No. M55534"
ORIGIN
Alignment Scores:
Pred. No.:      3,68e-31      Length:      1247
Score:          361.00      Matches:      67
Percent Similarity: 79.49%      Conservative: 26
Best Local Similarity: 57.26%      Mismatches:  18
Query Match:      56.49%      Indels:       6
DB:               6          Gaps:         2
US-10-657-740-1_COPY_51_173 (1-123) x AX401743 (1-1247)
Qy      7 LeuaspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuasp 26
Db      763 ATTGACACTGGCTCTCAGAGATGCGTATGGAGAGGACAGGTTCTCTGTAACCTGGAC 822
Qy      27 VallysHisPheSerProGluAspLeuThrVallysValGlnAspPheValGluIle 46
Db      823 GTGAAGCACTTCTCTCCAGAGAACTCAAAGTCAAGGTTCTGGGAGAGCGTATTGAGGTG 882
Qy      47 HisGlyIleHisAsnGluArgGlnAspHisGlyIleSerArgGluPheHisArg 66
Db      883 CACGCAAGCAGCAGAGAGCGCCAGCAGCAATGCTTCACTCCAGGAGTTCACACAGG 942
Qy      67 ArgTyArgLeuProSerAsnValaspGlnSerAlaLeuSerCysSerLeuSerAlaasp 86
Db      943 AAGTACCGATCCAGCAGCGAGTGGATCTCTCACCATTACTTCTCTGTCATGCGGCT 1002
Qy      87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106
Db      1003 GGAGTCTCTCACTGTGAATGGACCAAGAAACAG-----GCCTCTGGCCCTGAG 1050
Qy      107 ArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 121
Db      1051 CGCACCATTCCCATCCCGTGAAGAGAGCGCTGCTGTCTGTCATGCGAGCCCT 1101
RESULT 43
RATCRYAB
LOCUS
DEFINITION Rat alpha-crystallin B chain mRNA, complete cds.
ACCESSION M55534 J05699
VERSION M55534.1 GI:203609
KEYWORDS alpha-B-crystallin.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
  Rattus.
  1 (bases 1 to 1247)
  Iwaki,A., Iwaki,T., Goldman,J.E. and Liem,R.K.
  Multiple mRNAs of rat brain alpha-crystallin B chain result from
  alternative transcriptional initiation
  J. Biol. Chem. 265 (36), 22197-22203 (1990)
  91093055
  PUBMED 2176207
  COMMENT Original source text: Rat brain, cdna to mRNA, and liver DNA.
  FEATURES
    source
      Location/Qualifiers
        1..1247
          /organism="Rattus norvegicus"
          /mol_type="mRNA"
          /db_xref="taxon:10116"
          254..1247
            /gene="alpha (B) -crystallin"
          254..1247
            /gene="alpha (B) -crystallin"
            /product="alpha-crystallin B chain"
  gene
  mRNA
```

```
283..339
/gene="alpha (B) -crystallin"
/note="ORF1"
/codon_start=1
/protein_id="AAA40974.1"
/db_xref="GI:203610"
/translation="MGWVGATRCQRASETRP"
404..577
/gene="alpha (B) -crystallin"
/note="ORF2"
/codon_start=1
/protein_id="AAA40975.1"
/db_xref="GI:203611"
/translation="MCHPAKSLITSPHELSCSWNKPTDITVPEASQDCIYKQAVAA
AEGVRLTDTSLI"
441..530
/gene="alpha (B) -crystallin"
/note="ORF3"
/codon_start=1
/protein_id="AAA40976.1"
/db_xref="GI:203612"
/translation="MNCRGAGIIRKLTSPFOKLHKTAYIRGL"
583..1110
/gene="alpha (B) -crystallin"
/codon_start=1
/product="alpha-crystallin B chain"
/protein_id="AAA40977.1"
/db_xref="GI:203613"
/translation="MDIAIHPIWRRPFPFPHSPRLPDQFFGHEHLLSDLFSTSL
SPYLRPSPSLRPSLWDITGLSEMEKDRFVNLDVKHPSPEELKVKVLGDVIRVHG
KHREQRDEHGFISREPHKRYIPADVDPLITISLSLSDGLVTNGPRKQASGPRTIP
ITREKPAVTAAPKK"
ORIGIN
Alignment Scores:
Pred. No.:      3,68e-31      Length:      1247
Score:          361.00      Matches:      67
Percent Similarity: 79.49%      Conservative: 26
Best Local Similarity: 57.26%      Mismatches:  18
Query Match:      56.49%      Indels:       6
DB:               10         Gaps:         2
US-10-657-740-1_COPY_51_173 (1-123) x RATCRYAB (1-1247)
Qy      7 LeuaspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuasp 26
Db      763 ATTGACACTGGCTCTCAGAGATGCGTATGGAGAGGACAGGTTCTCTGTAACCTGGAC 822
Qy      27 VallysHisPheSerProGluAspLeuThrVallysValGlnAspPheValGluIle 46
Db      823 GTGAAGCACTTCTCTCCAGAGAACTCAAAGTCAAGGTTCTGGGAGAGCGTATTGAGGTG 882
Qy      47 HisGlyIleHisAsnGluArgGlnAspHisGlyIleSerArgGluPheHisArg 66
Db      883 CACGCAAGCAGCAGAGAGCGCCAGCAGCAATGCTTCACTCCAGGAGTTCACACAGG 942
Qy      67 ArgTyArgLeuProSerAsnValaspGlnSerAlaLeuSerCysSerLeuSerAlaasp 86
Db      943 AAGTACCGATCCAGCAGCGAGTGGATCTCTCACCATTACTTCTCTGTCATGCGGCT 1002
Qy      87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106
Db      1003 GGAGTCTCTCACTGTGAATGGACCAAGAAACAG-----GCCTCTGGCCCTGAG 1050
Qy      107 ArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 121
Db      1051 CGCACCATTCCCATCCCGTGAAGAGAGCGCTGCTGTCTGTCATGCGAGCCCT 1101
RESULT 44
OCCRYAB
LOCUS
DEFINITION O.cuniculus mRNA for alpha-B-crystallin.
ACCESSION X95383
```

```

X95383.1 GI:1177578
VERSION      alpha-B-crystallin; cryab gene.
KEYWORDS     Oryctolagus cuniculus (rabbit)
SOURCE       Oryctolagus cuniculus
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
REFERENCE    1
AUTHORS      Krausz, E., Augusteyn, R.C., Quinlan, R.A., Reddan, J.R., Russell, P.,
              Sax, C.M. and Graw, J.
TITLE        Expression of Crystallins, Pax6, Filensin, CP49, MIP, and MP20 in
              lens-derived cell lines
JOURNAL      Invest. Ophthalmol. Vis. Sci. 37 (10), 2120-2128 (1996)
MEDLINE      96409169
PUBMED       8814151
REFERENCE    2 (bases 1 to 548)
AUTHORS      Graw, J.
TITLE        Direct Submission
JOURNAL      Submitted (29-JAN-1996) J. Graw, Institute of Mammalian Genetics,
              GSF-Research Center Neuherberg, Ingolstaedter Landstrasse 1,
              Oberschleissheim, D-85764, FRG
FEATURES     Location/Qualifiers
              1..548
                /organism="Oryctolagus cuniculus"
                /mol_type="mRNA"
                /sub_species="New Zealand white"
                /db_xref="taxon:9986"
                /tissue_type="lens"
              1..548
                /gene="cryab"
              21..548
                /gene="cryab"
              /codon_start=1
              /product="alpha-B-crystallin"
              /protein_id="CAA64669.1"
              /db_xref="GI:1177579"
              /db_xref="GOA:P41316"
              /db_xref="UniProt/Swiss-Prot:P41316"
              /translation="MDIAIHHPWIRPFPFPHSPSLRFDQFGEHLLSDLPFTSTSL
              SPFLRPSPFLAPSWIDTGLSEMRLEKDRFSVNLVDKHFSPPELKVGLDGVIEVHG
              KHEERQDEHGFISREFHRYRIPADVDPITITSSSDGLVTNGPRKQAGPERTIP
              ITREKPAVTAAPKK"
ORIGIN
Alignment Scores:
Pred. No.:      3,086-31      Length:      548
Score:          358.00      Matches:      67
Percent Similarity: 78.63%      Conservative: 25
Best Local Similarity: 57.26%      Mismatches: 19
Query Match:     56.03%      Indels:      6
DB:              4          Gaps:      2

US-10-657-740-1_COPY_51_173 (1-123) x OCCRYAB (1-548)
QY      7  LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26
Db      201 ATTGACACTGGACTCTCAGATGCGCTGGAGAGGACAGTCTCTGTCAACCTGGAT 260
QY      27 ValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGluIle 46
Db      261 GTGAAGCACTTCTCCCGCAGAGAGCTCAAGGTCAAAGTTGGGTGATGTCATTGAGGTG 320
QY      47 HisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArg 66
Db      321 CACGCCAACATGAAGACGCCAGATGAACATGGTTTCATCTCCAGGAGTTCCACAGG 380
QY      67 ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86
Db      381 AAATACCGATCCCGAGTGTGGACCCCTCTACCATTTACTTCACTCATCTCATCTCAT 440
QY      87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106
Db      441 GGGGTCTCTCACTGTGAATGGACCAAGAACAA-----GCCCTCGGCCAGAG 488

```

---

```

107 ArgAlaIleProValSerArgGluGlyLysPro-----ThrSerAlaPro 121
Db      489 CGCACATTCCCATACCGTGAAGAGAAGCGTGTGTCTACTGTCAGCCCCC 539
RESULT 45
LOCUS      SEH272441          511 bp      mRNA      linear      ROD 26-FEB-2001
DEFINITION Spalax ehrenbergi partial mRNA for alphaB-crystallin (cryab gene).
ACCESSION  AJ272441
VERSION     AJ272441.1  GI:13162242
KEYWORDS   alphaB-crystallin; cryab gene
SOURCE     Nannospalax ehrenbergi (Ehrenberg's mole-rat)
ORGANISM   Nannospalax ehrenbergi
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Spalacinae;
            Nannospalax.
REFERENCE   1
AUTHORS     Smulders, R.H.P.H., van Dijk, M.M.A., Hoevenaars, S., Lindner, R.A.,
            Carver, J.A. and de Jong, W.W.
TITLE       The evolutionary fate of mole rat alpha-A-crystallin: a redundant
            but indispensable eye lens protein?
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 511)
AUTHORS     van Dijk, M.A.
TITLE       Direct Submission
JOURNAL     Submitted (28-FEB-2000) van Dijk M.A., Biochemistry, University of
            Nijmegen, Adelbertusplein 1, 6500 HB Nijmegen, NETHERLANDS
FEATURES     Location/Qualifiers
              1..511
                /organism="Nannospalax ehrenbergi"
                /mol_type="mRNA"
                /db_xref="taxon:30637"
              1..511
                /gene="cryab"
              1..511
                /gene="cryab"
                /function="shsp"
                /codon_start=1
                /product="alphaB-crystallin"
                /protein_id="CAC33095.1"
                /db_xref="GI:13162243"
                /db_xref="UniProt/TrEMBL:O9ND2"
                /translation="MDIAIHHPWIRPFPFPHSPSLRFDQFGEHLLSDLPFTSTSL
                SPFLRPSPFLAPSWIDTGLSEMRLEKDRFSVNLVDKHFSPPELKVGLDGVIEVHG
                KHEERQDEHGFISREFHRYRIPADVDPITITSSSDGLVTNGPRKQAGPERTIP
                ITREKPAVT"
ORIGIN
Alignment Scores:
Pred. No.:      3,686-31      Length:      511
Score:          357.00      Matches:      67
Percent Similarity: 78.15%      Conservative: 26
Best Local Similarity: 56.30%      Mismatches: 20
Query Match:     55.87%      Indels:      6
DB:              2          Gaps:      2

US-10-657-740-1_COPY_51_173 (1-123) x SEH272441 (1-511)
QY      1  SerLeuPheArg-----ThrValLeuAspSerGlyIleSerGluValArgSerAspArg 18
Db      157 TCCTTCTTTTCGGGCACCCAGCTGGATGACACTGGACTCTCAGAGATGCGTATGGAGAAG 216
QY      19 AspLysPheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrValLys 38
Db      217 GACAGTTCTCTGTCAACTGGATGGAAGCACTTCTCCCGCAGAGCAACTCAAGGTCAAG 276
QY      39 ValGlnAspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspHisGly 58
Db      277 GTGCTGGAGAGATGTCATTGAGTGCACGCAAGCATGAAGACGCGCAGGATGAGCATGCT 336
QY      59 TyrIleSerArgGluPheHisArgTyrArgLeuProSerAsnValAspGlnSerAla 78
Db      337 TTCTACTCCAGAGAGTTCCCGGAAGTACCGGATCCCGAGCTGATGTGGACCTCTCTACT 396

```







```

source
1. 528
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="GH00548X1.0"
/clone_lib="BD Creator (TM) CDS Library derived from MGC collection"
/lab_host="DH5alpha T1 resistant"
/note="Vector: pDNR-Dual"
1. 528
CDS
/product="crystallin, alpha B"
/protein_id="AAP35416.1"
/db_xref="GI:30582379"
/translating="MDIAIHHPWIRPPFPHPSPRLFDQFFGHEHLLSDLPFTSTL
SPFYLRPSFLRPSWFTDGLSEMRLEKDFSVNLDVKHFSPEELKVKVLGDVIEVHG
KHEERQDEHGFIREFHRYKIPADVDPLTITSSLSGDLTVNGPRKQVSGPERTIP
ITREKPAVTAAPKK"

ORIGIN
Alignment Scores:
Pred. No.: 8,37e-31 Length: 528
Score: 354.00 Matches: 66
Percent Similarity: 78.45% Conservat: 25
Best Local Similarity: 56.90% Mismatches: 19
Query Match: 55.40% Indels: 6
DB: 9 Gaps: 2

US-10-657-740-1_COPY_51_173 (1-123) x BT006770 (1-528)

Qy 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27
Db 184 GACACTGGACTCTCAGAGATCGCGCTGGAAAGGACAGGTTCTCTGTCAACCTGGATGTG 243
Qy 28 LysHisPheSerProGluAspLeuThrValIleValGlnAspAspPheValGluLeuHis 47
Db 244 AAGCACTTCTCCAGAGAACTCAAGTTAAGGTGTTGGAGATGTGATTGAGTGTGAT 303
Qy 48 GlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArg 67
Db 304 GGAACATGATGAAGAGCCGAGATGACATGTTTCACTCCAGGAGTTCACACAGGAAA 363
Qy 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87
Db 364 TACCGATCCAGCTGATGTAGACCTCTCACCATTACTTCTCATCTCTCATCTCATGG 423
Qy 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107
Db 424 GTCTCTCACTGTAATGACCAAGAAACAG-----GTCTCTGGCCCTGAGCGC 471
Qy 108 AlaIleProValSerArgGluLysPro-----ThrSerAlaPro 121
Db 472 ACCATTCCCATCCCGTGAAGAGAGAGCTGTCTGTACCCGAGCCCC 519

RESULT 50
BT007909
LOCUS
DEFINITION Synthetic construct Homo sapiens crystallin, alpha B mRNA, partial
cde.
BT007909 528 bp mRNA linear SYN 13-MAY-2003
VERSION BT007909.1 GI:30584656
KEYWORDS FLI CDNA.
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1 (bases 1 to 528)
AUTHORS Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
Phelan,M. and Farmer,A.
TITLE Cloning of human full-length CDSs in BD Creator (TM) System Donor
vector
JOURNAL unpublished
REFERENCE 2 (bases 1 to 528)

```

```

AUTHORS Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
Phelan,M. and Farmer,A.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
Circle, Palo Alto, CA 94303, USA
COMMENT This CDS clone is a part of a collection of human full length
expression clones generated by BD Biosciences Clontech and the
Harvard Institute of Proteomics. Each CDS has been cloned in two
forms: with and without stop-codon (to allow fusion with C-terminal
tag). The CDS has been directionally cloned using BD In-Fusion(TM)
cloning system between the SalI and HindIII sites of the pDNR-DUAL
vector. Additional sequences in the clone: 'ACC' after SalI site
and before 'ARG' to provide kozak consensus sequence; 'GG' after
last codon and before HindIII site to maintain reading frame.
Clone distribution: http://bioinfo.clontech.com/orfclones.
FEATURES
Location/Qualifiers
1..528
/organism="synthetic construct"
/mol_type="mRNA"
/db_xref="taxon:32630"
/clone="GH00548L1.0"
/clone_lib="BD Creator (TM) CDS Library derived from MGC collection"
/lab_host="DH5alpha T1 resistant"
/note="Vector: pDNR-Dual"
1..528
CDS
/note="Mutations: 527:Stop->Leu"
/codon_start=1
/translating="MDIAIHHPWIRPPFPHPSPRLFDQFFGHEHLLSDLPFTSTL
SPFYLRPSFLRPSWFTDGLSEMRLEKDFSVNLDVKHFSPEELKVKVLGDVIEVHG
KHEERQDEHGFIREFHRYKIPADVDPLTITSSLSGDLTVNGPRKQVSGPERTIP
ITREKPAVTAAPKKL"

ORIGIN
Alignment Scores:
Pred. No.: 8,37e-31 Length: 528
Score: 354.00 Matches: 66
Percent Similarity: 78.45% Conservat: 25
Best Local Similarity: 56.90% Mismatches: 19
Query Match: 55.40% Indels: 6
DB: 12 Gaps: 2

US-10-657-740-1_COPY_51_173 (1-123) x BT007909 (1-528)

Qy 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27
Db 184 GACACTGGACTCTCAGAGATCGCGCTGGAAAGGACAGGTTCTCTGTCAACCTGGATGTG 243
Qy 28 LysHisPheSerProGluAspLeuThrValIleValGlnAspAspPheValGluLeuHis 47
Db 244 AAGCACTTCTCCAGAGAACTCAAGTTAAGGTGTTGGAGATGTGATTGAGTGTGAT 303
Qy 48 GlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArg 67
Db 304 GGAACATGATGAAGAGCCGAGATGACATGTTTCACTCCAGGAGTTCACACAGGAAA 363
Qy 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87
Db 364 TACCGATCCAGCTGATGTAGACCTCTCACCATTACTTCTCATCTCTCATCTCATGG 423
Qy 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107
Db 424 GTCTCTCACTGTAATGACCAAGAAACAG-----GTCTCTGGCCCTGAGCGC 471
Qy 108 AlaIleProValSerArgGluLysPro-----ThrSerAlaPro 121
Db 472 ACCATTCCCATCCCGTGAAGAGAGAGCTGTCTGTACCCGAGCCCC 519

```

Tue May 31 10:29:17 2005

us-10-657-740-1\_copy\_51\_173.rge

Page 32

Search completed: May 30, 2005, 07:16:48  
Job time : 3109.55 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 29, 2005, 22:54:38 ; Search time 374.402 Seconds

(without alignments)  
1944.776 Million cell updates/sec

Title: US-10-657-740-1\_COPY\_51\_173

Perfect score: 639

Sequence: 1 SLFRVLDGISEVRSRDK.....HAERAIPVSRREKPTSAFSS 123

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO\_spool/US10657740/runat\_27052005\_165252\_3305/app\_query.fasta\_1.590  
-DB=N Geneseq 16Dec04 -QWTF=fastap -SUFFIX=ring -MINMATCH=0.1 -LOPCL=0  
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=150 -DOCLALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=50  
-MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10657740@cgn2\_1.1.796 @runat\_27052005\_165252\_3305 -NCPU=6 -ICPU=3  
-NO.MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DISPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 16Dec04:\*

1: geneseqm1980s:\*

2: geneseqm1990s:\*

3: geneseqm2000s:\*

4: geneseqm2001as:\*

5: geneseqm2001bs:\*

6: geneseqm2002as:\*

7: geneseqm2002bs:\*

8: geneseqm2003as:\*

9: geneseqm2003bs:\*

10: geneseqm2003cs:\*

11: geneseqm2003ds:\*

12: geneseqm2004as:\*

13: geneseqm2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	639	100.0	372	13	Adq78288 Human N-t
2	599	93.7	1056	10	AdB52521 Primary r
3	599	93.7	1056	10	AdF30549 Rat angio
4	577.5	90.4	1271	10	AdF30547 Rat angio
5	567	88.7	531	8	Abx12062 Human alp

6	367	57.4	211	12	ACH93386	Ach93386 Human gen
7	367	57.4	573	12	ACH79686	Ach79686 Human gen
8	361	56.5	528	10	ADB52642	ADB52642 Primary r
9	361	56.5	528	10	ABT41872	ABt41872 Toxicity
10	361	56.5	1247	6	ABK63512	ABk63512 Rat seque
11	361	56.5	1247	12	ADP72683	ADp72683 Renal tox
12	358	56.0	548	6	ABZ35138	ABz35138 Human gen
13	354	55.4	528	13	ADR46388	Adr46388 Human ten
14	354	55.4	537	10	ADE75374	AdE75374 Human mit
15	354	55.4	537	10	ADE75375	AdE75375 Human wil
16	354	55.4	691	2	AAx39668	Aax39668 Renal can
17	354	55.4	691	2	ABL65204	AbL65204 Lung canc
18	354	55.4	691	6	ABL62427	AbL62427 Colon ade
19	354	55.4	691	6	ABT10966	ABt10966 Human bre
20	354	55.4	691	6	ABK64671	ABk64671 Human ben
21	354	55.4	691	12	ADF09686	AdF09686 Human cry
22	354	55.4	691	10	ADN05800	AdN05800 Antipsori
23	354	55.4	691	13	ADR24825	Adr24825 Breast ca
24	354	55.4	856	3	AAC03893	Aac03893 Human sec
25	354	55.4	893	3	AAC10867	Aac10867 Human sec
26	354	55.4	911	3	AAC10863	Aac10863 Human sec
27	354	55.4	913	3	AAC10862	Aac10862 Human sec
28	354	55.4	927	3	AAC10861	Aac10861 Human sec
29	354	55.4	942	3	AAC03892	Aac03892 Human sec
30	354	55.4	1036	2	AAZ33574	Aaz33574 Human bre
31	332.5	52.0	380	8	ABX38978	ABx38978 Bovine ES
32	330	51.6	695	3	AAC03894	Aac03894 Human sec
33	283	44.3	449	8	ABX39112	ABx39112 Bovine ES
34	263	41.2	349	6	ABK87386	ABk87386 Target mo
35	261.5	40.9	615	12	ADN31647	Adn31647 Human hep
36	261.5	40.9	764	12	ADM94741	Adm94741 Human hep
37	261.5	40.9	789	11	ADI31974	Adi31974 Human cdn
38	261.5	40.9	847	10	ADG10693	Adg10693 Human STA
39	261.5	40.9	847	13	ADQ86169	AdQ86169 Human tum
40	261.5	40.9	865	6	ABO60780	ABo60780 Human HSB
41	261.5	40.9	865	13	ADR24765	Adr24765 Breast ca
42	261.5	40.9	865	13	ACN38656	Acn38656 Tumour-as
43	261.5	40.9	865	13	ADP23292	ADp23292 PRO polyo
44	261.5	40.9	1231	6	ABN97370	ABn97370 Gene #386
45	261.5	40.9	1231	6	ABN97370	ABn97370 Human hea
46	261.5	40.9	1380	3	AAA93441	AaA93441 GFP-HSP27
47	261.5	40.9	1380	6	ABS71560	ABs71560 DNA encod
48	261	40.8	604	6	ABK63074	ABk63074 Rat seque
49	261	40.8	604	12	ADP72259	ADp72259 Renal tox
50	260	40.7	599	6	ABV94670	ABv94670 Human pan
51	260	40.7	618	2	AAQ48718	AaQ48718 Encodes p
52	260	40.7	787	6	ABK63521	ABk63521 Rat seque
53	260	40.7	787	10	ADB58686	AdB58686 Toxicity-
54	260	40.7	787	10	ABT42261	ABt42261 Toxicity
55	260	40.7	787	12	ADP72725	ADp72725 Renal tox
56	260	40.7	787	13	ADT66606	Adt66606 Rat heat
57	260	40.7	1379	3	AAC66073	Aac66073 E. coli e
58	260	40.7	1379	4	AAF77805	Aaf77805 Fragment
59	260	40.7	1379	4	AAF61192	Aaf61192 Expressio
60	260	40.7	1379	5	AAD02211	Aad02211 Partial p
61	259	40.5	396	8	ABX47457	ABx47457 Bovine ES
62	258	40.4	719	4	ABL07843	ABl07843 Drosophi
63	256	40.1	560	10	ADK12087	Adk12087 Breast ca
64	250.5	39.2	480	10	ADH76210	Adh76210 Human hea
65	250.5	39.2	486	10	ADH76202	Adh76202 Rat heat
66	250.5	39.2	575	6	ABQ61117	ABq61117 Skeletal
67	250.5	39.2	1310	10	ABT41710	ABt41710 Toxicity
68	250.5	39.2	1433	5	ADM19230	Adm19230 Novel hum
69	250.5	39.2	1488	2	AzA2230	Aaz2230 Human nor
70	250.5	39.2	1506	5	ADM19501	Adm19501 Novel hum
71	250.5	39.2	1771	3	AAC76681	Aac76681 Human ORP
72	250.5	39.2	1820	12	ADQ22255	AdQ22255 Human sof
73	250	39.1	450	9	ACH17737	Ach17737 Human adu
74	248	38.8	503	6	ABL99507	ABl99507 Target ca
75	248	38.8	558	10	ADK12085	Adk12085 Breast ca
76	242.5	37.9	909	2	AAQ03969	AaQ03969 Clone P25
77	242	37.9	543	12	ADL11837	AdL11837 Cat flea
78	241.5	37.8	708	5	ADL40880	AdL40880 Human ova

c	79	240.5	37.6	376	4	AAL19387	Aa119387 Human bre
	80	240.5	37.6	411	11	ACN89210	Acn89210 Breast ca
	81	240.5	37.6	722	13	ADT66608	Adt66608 Murine hé
c	82	240.5	37.6	563	11	ACN87226	Acn87226 Breast ca
	83	234.5	36.7	614	5	ADI75644	Adi75644 Human ova
	84	234.5	36.7	614	5	ADI69298	Adi69298 Human ova
	85	233.5	36.5	755	2	AAQ03970	Aaq03970 Clone P25
	86	233.5	36.5	954	5	ADL62061	Adl62061 Human ova
	87	229	35.8	1700	4	AA556378	Aa556378 Human cDN
	88	226	35.4	421	9	ACH17560	Ach17560 Human adu
	89	223.5	35.0	500	9	ACH34922	Ach34922 Human end
	90	222	34.7	405	9	ACH16863	Ach16863 Human adu
c	91	216	33.8	2874	4	ABL07842	Ab107842 Drosophil
c	92	216	33.8	3119	4	ABL09418	Ab109418 Drosophil
c	93	212	33.2	49634	6	ABL68647	Ab168647 Kidney ca
	94	211.5	33.1	870	4	ABL07721	Ab107721 Drosophil
c	95	211.5	33.1	2870	4	ABL07720	Ab107720 Drosophil
	96	207	32.4	387	12	ADQ17563	Adq17563 Human sof
	97	202.5	31.7	937	4	ABL07315	Ab107315 Drosophil
	98	202.5	31.7	2651	4	ABL07714	Ab107714 Drosophil
	99	202.5	31.7	2937	4	ABL07314	Ab107314 Drosophil
	100	202	31.6	212	8	ABT33639	Abt33639 Anticance
c	101	198.5	31.1	396	6	ABL63575	Ab163575 Breast ca
c	102	198.5	31.1	396	6	ABL63995	Ab163995 Breast ca
c	103	197.5	30.9	549	3	AAc93846	Aac93846 Cat flea
	104	195	30.5	467	6	ABV94661	Abv94661 Human pan
	105	193	30.2	600	4	ABL07321	Ab107321 Drosophil
	106	193	30.2	2600	4	ABL07320	Ab107320 Drosophil
	107	192.5	30.1	482	9	ACH17221	Ach17221 Human adu
	108	192.5	30.1	550	12	ADL11752	Adl11752 Cat flea
	109	192	30.0	6806	9	ACF25379	Acf25379 Rat alpha
	110	188.5	29.5	1206	4	ABL07725	Ab107725 Drosophil
c	111	188.5	29.5	3206	4	ABL07724	Ab107724 Drosophil
	112	188	29.4	4206	6	ABZ35176	Abz35176 Human gen
	113	188	29.4	4206	10	ADE84874	Ade84874 Farnesyl
	114	182.5	28.6	393	3	AAc06428	Aac06428 Human sec
	115	182.5	28.6	410	6	ABs70907	Ab70907 Deer cDNA
c	116	179	28.0	218	4	AAc09000	Aac09000 Human bre
c	117	179	28.0	565	11	ACN79216	Acn79216 Breast ca
	118	174.5	27.3	1367	4	ABL07295	Ab107295 Drosophil
	119	174.5	27.3	3367	4	ABL07294	Ab107294 Drosophil
c	120	174	27.2	510	6	ABQ14062	Abq14062 Oligonuc1
	121	174	27.2	510	6	ABQ14063	Abq14063 Oligonuc1
	122	172	26.9	542	12	ADL11704	Adl11704 Cat flea
	123	170	26.6	654	5	AA542494	Aa542494 Human cDN
	124	165.5	25.9	487	13	ADS54436	Ads54436 Bacterial
	125	165.5	25.9	487	13	ADS54437	Ads54437 Bacterial
	126	163.5	25.6	1811	12	ADJ75836	Adj75836 Marker ge
	127	163	25.5	445	9	ACH23297	Ach23297 Human adu
	128	161.5	25.3	651	4	ABL07715	Ab107715 Drosophil
c	129	161.5	25.3	2651	4	ABL07714	Ab107714 Drosophil
c	130	161.5	25.3	2937	4	ABL07314	Ab107314 Drosophil
	131	160	25.0	1495	10	ADJ23459	Adj23459 Rat liver
	132	159	24.9	1627	2	AAZ21959	Aaz21959 cDNA enco
	133	159	24.9	1627	9	ACH04117	Ach04117 Human cDN
	134	159	24.9	1843	4	AAf74328	Aaf74328 Human H11
	135	159	24.9	1843	4	AAf74307	Aaf74307 Human H11
	136	159	24.9	1854	5	ABx71326	Abx71326 Human met
	137	159	24.9	2004	5	ADM19232	Adm19232 Novel hum
	138	159	24.9	2010	12	ADJ74941	Adj74941 Marker ge
	139	159	24.9	2010	13	ADR25621	Adr25621 Breast ca
c	140	159	24.9	2012	4	AAK52551	Aak52551 Human pol
	141	159	24.9	2036	10	ADJ22637	Adj22637 Human liv
	142	159	24.9	2048	4	AAK51567	Aak51567 Human pol
	143	158	24.7	764	6	ABX84243	Abx84243 Human cDN
	144	158	24.7	810	12	ADQ22105	Adq22105 Human sof
	145	154	24.1	823	2	AAZ24422	Aaz24422 Human bla
	146	151.5	23.7	1176	4	ABL24579	Ab124579 Drosophil
	147	150	23.5	438	11	ADN31653	Adn31653 Nematode
	148	149.5	23.4	333	3	AAc06427	Aac06427 Human sec
	149	148	23.2	403	8	ABX38783	Abx38783 Bovine ES
	150	148	23.2	490	13	ADS54435	Ads54435 Bacterial

ALIGNMENTS

RESULT 1

ADQ78288

ID ADQ78288 standard; cDNA; 372 BP.

XX

AC ADQ78288;

XX

XX 04-NOV-2004 (first entry)

XX

XX Human N-terminal truncated alpha-crystallin DNA.

XX

XX alpha-crystallin; enhanced protein expression;

KW enhanced protein secretion; protein aggregation; heat tolerance;

KW elevated temperature; human; ds; gene.

XX

OS Homo sapiens.

XX

XX Key Location/Qualifiers

XX CDS 1..372

XX

XX /\*tag= a

XX

XX /partial

XX

XX /product= "N-terminal truncated alpha-crystallin"

XX

XX /note= "No start codon given"

XX

XX US2004157289-A1.

XX

XX 12-AUG-2004.

XX

XX 08-SEP-2003; 2003US-00657740.

XX

XX 06-SEP-2002; 2002US-0408680P.

XX

XX (SALE/) SALERNO J C.

XX

XX (HANN/) HANNA M.

XX

XX (KORE/) KORETZ J F.

XX

XX (CRON/) CRONE D.

XX

XX (SMIT/) SMITH S M E.

XX

XX Salerno JC, Hanna M, Koretz JF, Crone D, Smith SME;

XX

XX WPI; 2004-580268/56.

XX

XX P-PSDB; ADQ78289.

XX

XX New truncated alpha-crystallin polypeptide derived from a wild-type alpha

XX

XX -crystallin protein, useful for enhancing protein (e.g. insulin or

XX

XX alcohol dehydrogenase) expression or secretion and for preventing protein

XX

XX aggregation.

XX

XX Claim 16; SEQ ID NO 2; 33pp; English.

XX

XX The invention relates to a truncated alpha-crystallin polypeptide derived

XX

XX from a wild-type alpha-crystallin protein, where the truncated

XX

XX polypeptide lacks an N-terminal sequence present in the wild-type

XX

XX protein. The composition and methods are useful for enhancing protein

XX

XX (e.g. insulin or alcohol dehydrogenase) expression or secretion and for

XX

XX preventing protein aggregation. These may also be used for creating a

XX

XX thermophilic host that tolerates elevated temperatures. The present

XX

XX sequence represents human N-terminal truncated alpha-crystallin DNA.

XX

XX

SQ Sequence 372 BP; 66 A; 136 C; 98 G; 72 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

Length:

Matches:

Conservative:

Mismatches:

Indels:

Gaps:

8.07e-75

639.00

100.00%

100.00%

100.00%

13

372

123

0

0

0

US-10-657-740-1\_COPY\_51\_173 (1-123) x ADQ78288 (1-372)



```

DE XX Rat angiogenesis modulating protein cdna #29.
KW ss; gene; rat; angiogenesis; angiogenesis modulating protein;
KW retinal neovascularisation; choroidal neovascularisation;
KW chronic inflammation; myocardial ischaemia; stroke;
KW coronary artery disease; peripheral vascular disease.
XX
OS Rattus norvegicus.
XX
XX US2003162706-A1.
XX
XX 28-AUG-2003.
XX
XX 10-DEC-2002; 2002US-00316253.
XX
XX 08-FEB-2002; 2002US-0355295P.
XX
XX 26-JUN-2002; 2002US-0391758P.
XX
XX (PROC ) PROCTER & GAMBLE CO.
XX
XX Peters KG, Thompson LJ, Wang F, Greis KD;
XX
XX WPI; 2003-711557/67.
XX
XX P-PSDB; ADF30550.
XX
XX Treating angiogenesis-mediated disorder, e.g., retinal or choroidal
XX neovascularization or diseases associated with chronic inflammation,
XX myocardial ischemia, stroke, coronary artery disease or peripheral
XX vascular disease.
XX
XX Disclosure; SEQ ID NO 112; 26pp; English.
XX
XX The invention relates to a method of treating an angiogenesis-mediated
XX disorder in a subject. The method is useful for treating angiogenesis-
XX mediated disorder, e.g., retinal or choroidal neovascularisation or
XX diseases associated with chronic inflammation, myocardial ischaemia,
XX stroke, coronary artery disease or peripheral vascular disease. The
XX present sequence is used in the exemplification of the invention.
XX
XX Sequence 1056 BP; 199 A; 333 C; 269 G; 255 T; 0 U; 0 Other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 6,99e-69 Length: 1056
XX Score: 599.00 Matches: 115
XX Percent Similarity: 97.56% Conservative: 5
XX Best Local Similarity: 93.50% Mismatches: 3
XX Query Match: 93.74% Indels: 0
XX DB: 10 Gaps: 0
XX
XX US-10-657-740-1_COPY_51_173 (1-123) x ADF30549 (1-1056)
XX
XX QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20
XX Db 163 TCTCTCTCCGACAGTGTGGACTCCGGCATCTCTGAGTCCGATCGACGGGACAG 222
XX
XX QY 21 PheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrValLysValGln 40
XX Db 223 TTGTGTCATCTCTGGATGTGAAGCACTTCTCTCTGAGGACCTCACCGTGAAGTACTG 282
XX
XX QY 41 AspAspPheValGluIleHisGlyLysHisGluArgGlnAspAspHisGlyTyrIle 60
XX Db 283 GAAGATTTCGTGGAGATCCATGGCAACACACAGAGGAGGATGACCATGGGTACATT 342
XX
XX QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80
XX Db 343 TCCCGTGAATTCACCGTCGCTACCGTCTCGCTTCCAAATGTGGACCACTCCGCTCTCC 402
XX
XX QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100
XX Db 403 TCTCTCTGTCTGCGGATGGCATGCTGACCTTCTCTGCGCCCAAGTCCAGTCTGGCTTG 462
XX
XX QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120
XX Db 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120
XX
XX
XX Alignment Scores:
XX Pred. No.: 6,34e-66 Length: 1271
XX Score: 577.50 Matches: 115
XX Percent Similarity: 82.19% Conservative: 5
XX Best Local Similarity: 78.77% Mismatches: 3
XX Query Match: 90.38% Indels: 23
XX DB: 10 Gaps: 1
XX
XX US-10-657-740-1_COPY_51_173 (1-123) x ADF30547 (1-1271)
XX
XX QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGlu----- 13
XX Db 309 TCTCTCTCCGACAGTGTGGACTCCGGCATCTCTGAGCTCATGACCCATATGTGTTT 368
XX
XX QY 14 -----ValArgSerAsp 17
XX Db 369 GTAATGCACCAACCACTGCTGAAACCCCAAGAACCAACCCCGCAAGGTCCGATCTGAC 428

```



QY 18 ArgAspLysPheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrVal 37  
 Db |||||  
 439 CGGCAAGTTGTCTATCTCTGTGATGTGAAGCACTTCTCTCTGAGGACCTCACCGTG 488  
 QY 38 LysValGlnAspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspHis 57  
 Db |||||  
 489 AAGGTACTGGAAGATTCTGTGAGATCCATGCGCAACACAGAGGAGGATGACCAT 548  
 QY 58 GlyTyrIleSerArgGluPheHisArgTyrArgLeuProSerAsnValAspGlnSer 77  
 Db |||||  
 549 GGTACTATTTCCCGTGAATTTCCACGTCGCTACCGTCTGCTTCCAATGTGACACGATCC 608  
 QY 78 AlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGln 97  
 Db |||||  
 609 GCCCTCTCTGCTCTTGTCTGGATGCGATGCTGACCTTCTTGGCCCCAAGGTCCAG 668  
 QY 98 ThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProValSerArgGluLysPro 117  
 Db |||||  
 669 TCTGCTTGGATCTGGCCACAGCGAGAGGCGCATTCCTGCTCACGGGAGGAGGAGGCC 728  
 QY 118 ThrSerAlaProSerSer 123  
 Db |||||  
 729 AGCTCGGACCCCTCGTCC 746

## RESULT 5

ABX12062

ID ABX12062 standard; DNA; 531 BP.

AC ABX12062;

DT 16-MAY-2003 (first entry)

XX Human alpha BNAC crystallin chimera DNA.

DE Human; ds; gene; chimera; alpha BNAC crystallin; protein shelf life;

KW protein aggregation; accessible hydrophobic region increase; mutant;

KW larger size oligomer formation; intersubunit interaction increase;

KW larger aggregate formation; larger porous oligomer formation;

KW increased ellipticity; less solvent accessible tryptophan;

KW increased chaperone-like activity; alpha A crystallin;

KW alpha B crystallin.

OS Homo sapiens.

OS Synthetic.

XX Key

FH Location/Qualifiers

FT 1..531

FT /tag= a

FT /product= "Alpha BNAC crystallin"

FT /partial

FT /note= "No stop codon given"

FT misc\_feature

FT 1..247

FT /tag= b

FT /note= "Derived from DNA sequence encoding N-terminal

FT sequence of alpha B crystallin"

FT 248..531

FT /tag= c

FT /note= "Derived from DNA sequence encoding N-terminal

FT sequence of alpha A crystallin"

FT FT

XX US2002177192-A1.

XX 28-NOV-2002.

XX 26-MAR-2002; 2002US-00105427.

XX 28-MAR-2001; 2001US-0279223P.

XX (KUMA/) KUMAR L V S.

XX (RAOC/) RAO C M.

XX Kumar LVS, Rao CM;

XX PI

XX

DR WPI; 2003-298776/29.

XX P-PSDB; ABG76084.

PT New chimera alpha BNAC nucleic acid, useful for preventing aggregation of

PT proteins and also for increasing shelf life of proteins of pharmaceutical

PT value.

XX Claim 1; Fig 9; 17pp; English.

XX The invention relates to a chimera alpha BNAC polynucleotide that encodes

CC a chimeric alpha BNAC polypeptide. The polypeptide is also useful for

CC preventing protein aggregation. The polypeptide is also useful for

CC increasing the shelf life of proteins of pharmaceutical value. The

CC polypeptide shows an increase in accessible hydrophobic regions, forms

CC larger size oligomers, shows an increase in intersubunit interaction,

CC forms larger aggregates, forms larger porous oligomers and shows

CC increased ellipticity as compared to eye lens crystallins alpha A and

CC alpha B. The tryptophan residues in the polypeptide are less solvent

CC accessible as compared to those of eye lens crystallins alpha A and alpha

CC B. The polypeptide shows extraordinarily high chaperone-like activity

CC ranging between 3 - 6 times that of the eye lens crystallins alpha A and

CC alpha B. The present sequence represents the chimera DNA that encodes

CC human alpha BNAC crystallin

XX SQ Sequence 531 BP; 90 A; 198 C; 130 G; 113 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 4.41e-65 Length: 531

Score: 567.00 Matches: 106

Percent Similarity: 97.41% Conservative: 7

Best Local Similarity: 91.38% Mismatches: 3

Query Match: 88.73% Indels: 0

DB: 8 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x ABX12062 (1-531)

QY 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27

Db 184 GACACTGGACCTCTCAGAGATCGCCTGGAGAGGACAGGTTCTCTCAACCTGGATGTG 243

QY 28 LysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHis 47

Db 244 AAGCATTCTCCCGAGGAGCTCACCCTGAAGGTGACGACGACTTTGTGGAGATCCAC 303

QY 48 GlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArgArg 67

Db 304 GAAAGCACACAGAGCGCCAGACGACGACGCTACATTTCCCGTAGTTCCACCGCCGC 363

QY 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87

Db 364 TACGCGCTGCGCTCCACACGTCGACGACGTCGCGCCCTCTCTGCTCCCTGTCTGCCGATGCG 423

QY 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107

Db 424 ATGCTGACCTTCTGTGGCCCCCAAGATCCAGACTGCGCTGGATGCCACCCACGCGCGGA 483

QY 108 AlaIleProValSerArgGluLysProThrSerAlaProSerSer 123

Db 484 GCCATCCCGGTGTGCGGGAGGAGAGCCACCTCGGCTCCCTCGTCC 531

## RESULT 6

ACH93386

ID ACH93386 standard; DNA; 211 BP.

XX ACH93386;

AC ACH93386;

XX 29-JUL-2004 (first entry)

XX Human genome derived single exon probe #26581.

XX Human; probe; ss; gene expression; single exon probe; microarray;

XX alternative splicing event; genomic alteration.

XX

OS Homo sapiens.  
 XX US2003194704-A1.  
 XX 16-OCT-2003.  
 XX  
 XX 03-APR-2002; 2002US-00029386.  
 XX 03-APR-2002; 2002US-00029386.  
 XX (PENN/) PENN S G.  
 XX (RANK/) RANK D R.  
 XX (HANZ/) HANZEL D K.  
 XX  
 XX Penn SG, Rank DR, Hanzel DK;  
 XX WPI; 2004-119264/12.  
 XX  
 XX New human genome-derived single exon nucleic acid probes useful for human  
 XX gene expression analysis, for identifying or characterizing alternative  
 XX splicing events, for assessing genomic alterations or as tools for  
 XX surveying tissues.  
 XX  
 XX Claim 1; SEQ ID NO 26581; 80pp; English.  
 XX  
 XX The invention relates to a nucleic acid probe for measuring human gene  
 XX expression, comprising any of the 27,400 fully defined nucleotide  
 XX sequences in the specification, or their complements or fragments, and  
 XX encoding at least 8 amino acids of any of the 6888 amino acid sequences  
 XX fully defined in the specification. The probe is a single exon probe that  
 XX hybridises under high stringency conditions to a nucleic acid molecule  
 XX expressed in human cells or tissues. Also included are a spatially-  
 XX addressable set of single exon nucleic acid probes for measuring human  
 XX gene expression (comprising a plurality of single exon nucleic acid  
 XX probes cited above, where each of the plurality of probes is separately  
 XX and addressably isolatable or amplifiable from the plurality), a single  
 XX exon microarray for measuring human gene expression, a method of  
 XX measuring human gene expression, a vector comprising the single exon  
 XX probe cited above, an ORF-encoded peptide comprising at least 8  
 XX contiguous amino acids of any of the above-mentioned amino acid  
 XX sequences (optionally with conservative amino acid substitutions), an  
 XX isolated antibody that binds specifically to a peptide cited above,  
 XX methods of selling and/or licensing single exon probes or microarrays to  
 XX a customer desiring to measure gene expression, a method of providing  
 XX storage medium which contains a database having a plurality of records  
 XX (each record including data on the expression of a single exon probe  
 XX cited above. The probe, methods and apparatus are useful in gene  
 XX expression analysis. The probes may be used as tools for surveying  
 XX tissues to detect the presence of expressed messages that contain their  
 XX specific exon, or in constructing genome-derived single exon microarrays.  
 XX In addition, the probes are used in identifying and characterizing  
 XX alternative splicing events, in detecting and characterising gross  
 XX alterations in the genomic locus that includes their exon, in assessing  
 XX smaller genomic alterations, in priming the synthesis of nucleic acids,  
 XX or in expressing the ORF-encoded peptide. The present sequence is a human  
 XX single exon probe of the invention. Note: The sequence data for this  
 XX patent did not form part of the printed specification, but was obtained  
 XX in electronic format directly from USPTO at  
 XX seqdata.uspto.gov/sequence.html?docID=20030194704  
 XX  
 XX SQ Sequence 211 BP; 32 A; 85 C; 55 G; 39 T; 0 U; 0 Other;  
 XX  
 XX Alignment Scores:  
 XX Pred. No.: 3,156-39 Length: 211  
 XX Score: 367.00 Matches: 70  
 XX Percent Similarity: 100.00% Conservative: 0  
 XX Best Local Similarity: 100.00% Mismatches: 0  
 XX Query Match: 57.43% Indels: 0  
 XX DB: 12 Gaps: 0  
 XX  
 XX US-10-657-740-1\_COPY\_51\_173 (1-123) x ACH93386 (1-211)

QY 54 GlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgGlyTyrArgLeuProSerIle 73  
 Db 1 CAGGACGACACCGCTACATTTCCCGTGAGTTCCACCGCGCTACCGCTGCGTCCAAC 60  
 QY 74 ValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGly 93  
 Db 61 GTGGACCAAGTCGGCCCTCTCTTGTCTCCCTGTGCGATGCGATGCGATGCGATGCG 120  
 QY 94 ProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProValSerArg 113  
 Db 121 CCCAAGATCCAGACTGGCTGGATGCCACCCACCGCGAGCGAGCCATCCCGTGTGGCG 180  
 QY 114 GluGluLysProThrSerAlaProSerSer 123  
 Db 181 CAGGAGAAGCCACCTCGGCTCCCTCGTCC 210  
 RESULT 7  
 ACH79686  
 ID ACH79686 standard; DNA; 573 BP.  
 XX ACH79686;  
 XX  
 XX 29-JUL-2004 (first entry)  
 XX  
 XX Human genome derived single exon probe #12881.  
 XX Human; probe; ss; gene expression; single exon probe; microarray;  
 XX alternative splicing event; genomic alteration.  
 XX  
 XX Homo sapiens.  
 XX  
 XX US2003194704-A1.  
 XX 16-OCT-2003.  
 XX 03-APR-2002; 2002US-00029386.  
 XX 03-APR-2002; 2002US-00029386.  
 XX (PENN/) PENN S G.  
 XX (RANK/) RANK D R.  
 XX (HANZ/) HANZEL D K.  
 XX  
 XX Penn SG, Rank DR, Hanzel DK;  
 XX WPI; 2004-119264/12.  
 XX  
 XX New human genome-derived single exon nucleic acid probes useful for human  
 XX gene expression analysis, for identifying or characterizing alternative  
 XX splicing events, for assessing genomic alterations or as tools for  
 XX surveying tissues.  
 XX  
 XX Claim 15; SEQ ID NO 12881; 80pp; English.  
 XX  
 XX The invention relates to a nucleic acid probe for measuring human gene  
 XX expression, comprising any of the 27,400 fully defined nucleotide  
 XX sequences in the specification, or their complements or fragments, and  
 XX encoding at least 8 amino acids of any of the 6888 amino acid sequences  
 XX fully defined in the specification. The probe is a single exon probe that  
 XX hybridises under high stringency conditions to a nucleic acid molecule  
 XX expressed in human cells or tissues. Also included are a spatially-  
 XX addressable set of single exon nucleic acid probes for measuring human  
 XX gene expression (comprising a plurality of single exon nucleic acid  
 XX probes cited above, where each of the plurality of probes is separately  
 XX and addressably isolatable or amplifiable from the plurality), a single  
 XX exon microarray for measuring human gene expression, a method of  
 XX measuring human gene expression, a vector comprising the plurality of  
 XX probes cited above, an ORF-encoded peptide comprising at least 8  
 XX contiguous amino acids of any of the above-mentioned amino acid  
 XX sequences (optionally with conservative amino acid substitutions), an  
 XX isolated antibody that binds specifically to a peptide cited above,  
 XX methods of selling and/or licensing single exon probes or microarrays to  
 XX a customer desiring to measure gene expression, a method of providing  
 XX storage medium which contains a database having a plurality of records  
 XX (each record including data on the expression of a single exon probe  
 XX cited above. The probe, methods and apparatus are useful in gene  
 XX expression analysis. The probes may be used as tools for surveying  
 XX tissues to detect the presence of expressed messages that contain their  
 XX specific exon, or in constructing genome-derived single exon microarrays.  
 XX In addition, the probes are used in identifying and characterizing  
 XX alternative splicing events, in detecting and characterising gross  
 XX alterations in the genomic locus that includes their exon, in assessing  
 XX smaller genomic alterations, in priming the synthesis of nucleic acids,  
 XX or in expressing the ORF-encoded peptide. The present sequence is a human  
 XX single exon probe of the invention. Note: The sequence data for this  
 XX patent did not form part of the printed specification, but was obtained  
 XX in electronic format directly from USPTO at  
 XX seqdata.uspto.gov/sequence.html?docID=20030194704  
 XX  
 XX SQ Sequence 211 BP; 32 A; 85 C; 55 G; 39 T; 0 U; 0 Other;  
 XX  
 XX Alignment Scores:  
 XX Pred. No.: 3,156-39 Length: 211  
 XX Score: 367.00 Matches: 70  
 XX Percent Similarity: 100.00% Conservative: 0  
 XX Best Local Similarity: 100.00% Mismatches: 0  
 XX Query Match: 57.43% Indels: 0  
 XX DB: 12 Gaps: 0  
 XX  
 XX US-10-657-740-1\_COPY\_51\_173 (1-123) x ACH93386 (1-211)

CC human gene expression data by subscription, and a computer-readable  
 CC storage medium which contains a database having a plurality of records  
 CC (each record including data on the expression of a single exon probe  
 CC cited above). The probe, methods and apparatus are useful in gene  
 CC expression analysis. The probes may be used as tools for surveying  
 CC tissues to detect the presence of expressed messages that contain their  
 CC specific exon, or in constructing genome-derived single exon microarrays.  
 CC In addition, the probes are used in identifying and characterizing  
 CC alternative splicing events, in detecting and characterizing gross  
 CC alterations in the genomic locus that includes their exon, in assessing  
 CC smaller genomic alterations, in priming the synthesis of nucleic acids,  
 CC or in expressing the ORF-encoded peptide. The present sequence is a human  
 CC single exon probe of the invention. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from USPTO at  
 CC seqdata.uspto.gov/sequence.html?DocID=20030194704

XX SQ Sequence 573 BP; 86 A; 202 C; 171 G; 114 T; 0 U; 0 Other;

#### Alignment Scores:

\* Pred. No.: 1.33e-38 Length: 573  
 Score: 367.00 Matches: 70  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 57.43% Indels: 0  
 DB: 12 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x ACH79686 (1-573)

Qy 54 GlnAspSerHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeuProSerArg 73  
 Db 203 CAGGACGACACCGGTACATTTCCGTGAGTTCCACCCCGCTACCGCTCCGTCACAC 262  
 Qy 74 ValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGly 93  
 Db 263 GTGGACCACTCGGCCCTCTTGCTCCCTGCTGCGGATGGCATGCTGACCTTGTGGC 322  
 Qy 94 ProTylleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProValSerArg 113  
 Db 323 CCCAAGATCCAGACTGGCTGGATGCCACCCACGCGGAGCGGATCCCGTGTGCGGG 382  
 Qy 114 GluGluLysProThrSerAlaProSerSer 123  
 Db 383 GAGGAGAGCCCACTCGGCTCCCTCGTCC 412

RESULT 8  
 ADB52642  
 ID ADB52642 standard; DNA; 528 BP.  
 XX AC ADB52642;  
 XX DT  
 XX 04-DEC-2003 (first entry)  
 XX DE Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:3184.  
 XX KW toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;  
 XX KW toxicity marker; toxicity progression; drug screening;  
 XX KW primary rat hepatocyte toxicity modelling; gene; ds.  
 XX OS Rattus norvegicus.  
 XX FN WO2003065993-A2.  
 XX PD 14-AUG-2003.  
 XX PF 04-FEB-2003; 2003WO-US003482.  
 XX PR 04-FEB-2002; 2002US-0353171P.  
 XX PR 13-MAR-2002; 2002US-0363534P.  
 XX PR 08-APR-2002; 2002US-0370248P.  
 XX PR 10-APR-2002; 2002US-0371134P.  
 XX PR 10-APR-2002; 2002US-0371135P.  
 XX PR 10-APR-2002; 2002US-0371150P.

PR 11-APR-2002; 2002US-0371413P.  
 PR 19-APR-2002; 2002US-0373601P.  
 PR 19-APR-2002; 2002US-0373602P.  
 PR 22-APR-2002; 2002US-0374139P.  
 PR 08-MAY-2002; 2002US-0378370P.  
 PR 09-MAY-2002; 2002US-0378652P.  
 PR 09-MAY-2002; 2002US-0378653P.  
 PR 09-MAY-2002; 2002US-0378655P.  
 PR 09-JUL-2002; 2002US-0394230P.  
 PR 09-JUL-2002; 2002US-0394253P.  
 PR 04-SEP-2002; 2002US-0407688P.  
 PR 28-JAN-2003; 2003US-0442900P.

XX (GENE-) GENE LOGIC INC.

XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Orr M;  
 PI Elashoff M;

XX WPI; 2003-731472/69.

XX Determining if a compound induces a toxic effect on a tissue or cell, for  
 PT identifying hepatotoxic compounds, comprises comparing a gene expression  
 PT profile of a tissue or cell sample to a database of Tox mean and non-Tox  
 PT mean values.

XX Claim 44; SEQ ID NO 3184; 874pp; English.

XX The present invention describes a method for determining whether a  
 CC compound induces a toxic effect on a tissue or cell. The method comprises  
 CC preparing a gene expression profile of a tissue or cell sample exposed to  
 CC the compound, and comparing the gene expression profile to a database  
 CC comprising data or information on the Tox mean and non-Tox mean value.  
 CC The method is useful for predicting or identifying at least one toxic  
 CC effect, particularly hepatotoxicity, of a test or unknown compound. The  
 CC genes listed in the specification are useful as diagnostic or toxicity  
 CC markers for the prediction or identification of the physiological state  
 CC of tissue or cell sample that has been exposed to a compound, or to  
 CC identify or predict the toxic effects of a compound or an agent. These  
 CC may also be used as markers for monitoring toxicity progression or for  
 CC drug screening. The present sequence represents a primary rat hepatocyte  
 CC toxicity modelling related gene sequence from the present invention.

SQ Sequence 528 BP; 114 A; 170 C; 128 G; 116 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.: 7.32e-38 Length: 528  
 Score: 361.00 Matches: 67  
 Percent Similarity: 79.49% Conservative: 26  
 Best Local Similarity: 57.26% Mismatches: 18  
 Query Match: 56.49% Indels: 6  
 DB: 10 Gaps: 2

US-10-657-740-1\_COPY\_51\_173 (1-123) x ADB52642 (1-528)

Qy 7 LeuAspSerGlyIleSerGluValArgSerAspArgLysPheValIlePheLeuAsp 26  
 Db 181 ATTGACACTGGCTCTCAGAGATCGTATGGAGAGGACAGGTTCTCTGTGAACCTGGAC 240  
 Qy 27 ValIleHisPheSerProGluAspLeuThrValIleValGlnAspPheValGluIle 46  
 Db 241 GTGAAGCAGCTTCTCTCCAGAGAACTCAAGTCAAGGTTCTGGGAGAGCTGATGAGGTG 300  
 Qy 47 HisGlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArg 66  
 Db 301 CACGGCAGACGACGAGAGCGCAGGACGAACATGGTTCATCTCCAGGGAGTTCACAGG 360  
 Qy 67 ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86  
 Db 361 AAGTACCGGATCCAGCGGACGCTGCTCTCACCATTACTTCTTCTCTGTCATCGGAT 420  
 Qy 87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106  
 Db 421 GGAGTCTCTACTGTGAATGGACCAAGGAACAG-----GCCTCTGCGCCTGAG 468



PA (GENE-) GENE LOGIC INC.  
XX Mendrick D, Porter MW, Johnson KR, Castle AL, Elashoff MR;  
PI WPI; 2002-241625/29.  
XX Predicting toxic effects of compounds or the progression of these toxic  
XX effects by determining the changes in gene expression in tissues or cells  
XX exposed to the toxin and comparing these to gene expression in unexposed  
XX tissues or cells.  
XX Claim 1; SEQ ID NO 1419; 239pp; English.  
XX The invention relates to methods for predicting toxic effects of  
XX compounds or the progression of these toxic effects by determining the  
XX global changes in gene expression in tissues or cells exposed to the  
XX toxin and comparing these to gene expression in unexposed tissues or  
XX cells. Also included are methods of predicting at least one toxic effect  
XX of a compound or progression of a toxic effect, preferably the  
XX hepatotoxicity of a compound, comprising detecting the level of  
XX expression in a tissue or cell sample exposed to the compound of two or  
XX more genes listed in the specification, where differential expression of  
XX the genes is indicative of at least one toxic effect or progression. The  
XX method can also be used to identify an agent which modulates the toxic  
XX response and predict cellular pathways that a compound modulates in a  
XX cell. The methods utilize a set of at least two probes (on a solid  
XX support in kit form), where each of the probes comprises a sequence that  
XX specifically hybridises to a gene listed in the specification, a computer  
XX system comprising a database containing information identifying the  
XX expression level in a tissue or cell sample exposed to a hepatotoxin of a  
XX set of genes comprising at least two genes listed in the specification,  
XX and a user interface to view the information used to present information,  
XX identifying the expression level in a tissue or cell of at least one gene  
XX listed in the specification. The method is useful for elucidating global  
XX changes in gene expression and for identifying toxicity markers in  
XX tissues or cell exposed to a known toxin. The genes may be used as  
XX toxicity markers in drug screening and toxicity assays. The genes and  
XX gene expression information may be used as diagnostic markers for the  
XX prediction or identification of the physiological state of tissue or cell  
XX sample that has been exposed to a compound or agent. Hepatotoxicity is  
XX characterised by centrilobular necrosis and steatosis. The present  
XX sequence is an expressed sequence tag (EST) or cDNA derived from a gene  
XX which is differentially expressed in response to a hepatotoxic agent  
XX  
SQ Sequence 1247 BP; 294 A; 360 C; 306 G; 287 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 2,528-37 Length: 1247  
Score: 361.00 Matches: 67  
Percent Similarity: 79.49% Conservative: 26  
Best Local Similarity: 57.26% Mismatches: 18  
Query Match: 56.49% Indels: 6  
Gaps: 2  
US-10-657-740-1\_COPY\_51\_173 (1-123) x ABK63512 (1-1247)

QY 7 LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26  
Db ATTGACACTGGCTCTCAGAGATGGTATGGAGAGGACAGCTTCTGTGACCTGGAC 822

QY 27 ValLysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGluLe 46  
Db GTGAAGCACTTCTCTCCAGAGAACTCAAGGTCAGGAGCTGTGATGAGGTG 882

QY 47 HisGlyLysHisAsnGluArgGlnAspAspHisGlyIleSerArgGluPheHisArg 66  
Db CACGCGACAGCAGCAAGAGCGCCAGACCAACATGCTTCATCTCAGGAGGTTCACAGG 942

QY 67 ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86  
Db AAGTACCGGATCCAGCGGACGATGATCTCTCACCATTACTTCTCTCTCTCTCTCTCT 1002

QY 87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106

Db 1003 GGAGTCTCTACTGTGAATGGACCAAGGAACAG-----GCCTTGGCCCTGAG 1050  
QY 107 ArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 121  
Db 1051 CGCACCATTCCCATCACCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1101  
RESULT 11  
ADP72683  
ID ADP72683 standard; DNA; 1247 BP.  
XX AC ADP72683;  
XX 26-AUG-2004 (first entry)  
XX Renal toxin progression gene marker #1272.  
XX ds; toxic effect; gene expression profile; kidney tissue;  
XX differential gene expression; toxicity progression; toxicity marker;  
XX drug screening; toxicity assay; kidney pathology; nephritis;  
XX kidney necrosis; glomerular injury; tubular injury;  
XX focal segmental glomerulosclerosis.  
XX Rattus norvegicus.  
XX WO2004048598-A2.  
XX 10-JUN-2004.  
XX 24-NOV-2003; 2003WO-US037556.  
XX 22-NOV-2002; 2002US-00301856.  
XX (GENE-) GENE LOGIC INC.  
XX Mendrick DL, Porter MW, Johnson KR, Castle A, Higgs B;  
XX Elashoff M;  
XX WPI; 2004-460771/43.  
XX Predicting (the progression of) a toxic effect of a compound, for  
XX monitoring the progression of renal disease states, comprises preparing a  
XX gene expression profile of a kidney tissue or cell sample exposed to the  
XX compound.  
XX Claim 11; SEQ ID NO 1272; 266pp; English.  
XX The invention relates to a method of predicting (the progression of) a  
XX toxic effect of a compound by preparing a gene expression profile of a  
XX kidney tissue or cell sample exposed to the compound and comparing the  
XX gene expression profile to a database, or detecting the level of gene(s)  
XX expression in a tissue or cell sample exposed to the compound, where  
XX differential gene expression compared to a control indicates a toxic  
XX effect (toxicity progression). The method is useful for predicting (the  
XX progression of) at least one toxic effect of a compound. The genes are  
XX useful as toxicity markers in drug screening and toxicity assays. The  
XX methods are useful for predicting the likelihood that a compound or test  
XX agent will induce various specific kidney pathologies, such as nephritis,  
XX kidney necrosis, glomerular and tubular injury, or focal segmental  
XX glomerulosclerosis. The methods are useful for determining the similarity  
XX of a toxic response to one or more individual compounds and for  
XX predicting or elucidating the potential cellular pathways influenced,  
XX induced or modulated by the compound or test agent. The kit is useful for  
XX predicting or modelling the toxic response of a test compound, for  
XX monitoring the progression of renal disease states, for identifying genes  
XX that show promise as new drug targets and for screening known and newly  
XX designed drugs. This sequence corresponds to a gene marker used in the  
XX method of the invention. (Note: The sequence data for this patent did not  
XX form part of the printed specification, but was obtained in electronic  
XX format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences).  
XX Sequence 1247 BP; 294 A; 360 C; 306 G; 287 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 2.52e-37 Length: 1247  
 Score: 361.00 Matches: 67  
 Percent Similarity: 79.49% Conservative: 26  
 Best Local Similarity: 57.26% Mismatches: 18  
 Query Match: 56.49% Indels: 6  
 DB: 12 Gaps: 2

US-10-657-740-1\_COPY\_51\_173 (1-123) x ADP72693 (1-1247)

QY 7 LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26  
 DB 763 ATTGACACTGGCTCTCAGAGATCGGTATGAGAGACACAGGTTCTCTGGAACCTGGAC 822  
 QY 27 VallyshHisPheSerProGluAspLeuThrVallyshValGlnAspPheValGluIle 46  
 DB 823 GTGAAGCACTTCTCCAGAGGAACCTCAAAGTCAAGGTCTGGAGACGTGATGAGGTG 882  
 QY 47 HisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArg 66  
 DB 883 CACGGCAAGCAGCAAGAGCGCCAGGACGAACATGGCTTCATCTCCAGGGAGTTCCACAGG 942  
 QY 67 ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86  
 DB 943 AAGTACCGGATCCAGCCGCGTGGATCTCTCACCATTACTTCTTCCCTGTCATCGGAT 1002  
 QY 87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106  
 DB 1003 GGAGTCTCTACTGTGAATGACCAAGGAACAG-----GCCCTGGCCCTGAG 1050  
 QY 107 ArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 121  
 DB 1051 CGCACCATTCCTCCATCCCGGTGAAGAGAGACCTGCTGCTCACTGCAGCCCT 1101

## RESULT 12

ABZ35138

ID ABZ35138 standard; cDNA; 548 BP.

AC ABZ35138;

XX 05-FEB-2003 (first entry)

XX Human gene expression profile polynucleotide SEQ ID NO 250.

XX Human; artery; endothelium; umbilical; vein; aorta; pulmonary artery;  
 KW bronchial epithelium; prostate; muscle; lung fibroblast; osteoblast;  
 KW tumour; microarray; genome mapping; antibiotic; antiviral; antifungal;  
 KW gene expression; gene; ss.

XX Homo sapiens.

OS WO200274979-A2.

FN 26-SEP-2002.

XX 20-MAR-2002; 2002WO-US008456.

XX 20-MAR-2001; 2001US-0276947P.

PR (ORTH ) ORTHO CLINICAL DIAGNOSTICS INC.

XX Wan J, Wang Y;

XX WPI; 2002-740862/80.

XX New gene expression profile generated from primary, endothelial,  
 PT epithelial, and muscle cell types, useful for identifying disease  
 PT pathologies involving alterations of gene expression, e.g. cancer.  
 XX Claim 10; Page 429; 850pp; English.

XX The invention relates to a gene expression profile comprising one or more  
 CC genes (ABZ34889-ABZ35692) and generated from a cell type. The cell type

CC is a coronary artery endothelium, umbilical artery or vein endothelium,  
 CC aortic endothelium, dermal microvascular endothelium, pulmonary artery  
 CC endothelium, myometrium microvascular endothelium, keratinocyte  
 CC epithelium, bronchial epithelium, mammary epithelium, prostate  
 CC epithelium, renal cortical epithelium, renal proximal tubule epithelium,  
 CC small airway epithelium, renal epithelium, umbilical artery smooth  
 CC muscle, neonatal dermal fibroblast, pulmonary artery smooth muscle,  
 CC dermal fibroblast, neural progenitor cells, skeletal muscle, astrocytes,  
 CC aortic smooth muscle, mesangial cells, coronary artery smooth muscle,  
 CC bronchial smooth muscle, uterine smooth muscle, lung fibroblast,  
 CC osteoblasts or prostate stromal cell. The gene expression profile is used  
 CC for determining the level of RNA expression for a sample, determining the  
 CC phenotype of a cell and distinguishing cell types. The gene or a protein  
 CC expression profile is useful in identifying disease pathologies involving  
 CC alterations of gene expression. The assessment of expression profiles may  
 CC provide meaningful information with respect to tumour type and stage,  
 CC treatment methods, and prognosis. The gene or protein expression profile  
 CC may also be used for creating microarrays. The microarray is useful for  
 CC genetic and physical mapping of genomes, DNA sequencing, genetic or  
 CC medical diagnosis, genotyping of organisms, confirming cell or tissue  
 CC identifications and in identifying promising antibiotics, antiviral or  
 CC antifungal agents

SQ Sequence 548 BP; 120 A; 178 C; 127 G; 123 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 1.92e-37 Length: 548  
 Score: 358.00 Matches: 67  
 Percent Similarity: 78.63% Conservative: 25  
 Best Local Similarity: 57.26% Mismatches: 19  
 Query Match: 56.03% Indels: 6  
 DB: 6 Gaps: 2

US-10-657-740-1\_COPY\_51\_173 (1-123) x ABZ35138 (1-548)

QY 7 LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26  
 DB 201 ATTGACACTGGCTCTCAGAGATCGCTCGGAGAGGACAGGTTCTCTGCAACCTGGAT 260

QY 27 VallyshHisPheSerProGluAspLeuThrVallyshValGlnAspPheValGluIle 46  
 DB 261 GTGAAGCACTTCTCCCGAGAGGAGTCAAGTCAAGTGTGGTGTGATGTGATGAGGTG 320

QY 47 HisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArg 66  
 DB 321 CACGGCAAGCAGCAAGAGCGCCAGGATGAACATGGTTTCATCTCCAGGGAGTTCCACAGG 380

QY 67 ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86  
 DB 381 AAATACCGGATCCAGCTGATGTGGACCTCTCACCATTACTTCTCCCTGTCATCTGAT 440

QY 87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106  
 DB 441 GGGGTCTCTCACTGTGAATGACCAAGGAACAA-----GCCCTGGCCCGAG 488

QY 107 ArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 121

DB 489 CGCACCATTCCTCCATCCCGGTGAAGAGAGCCTGCTGCTCACTGCAGCCCT 539

## RESULT 13

ADR46388

ID ADR46388 standard; DNA; 528 BP.

XX ADR46388;

XX 04-NOV-2004 (first entry)

DT Human lens crystalline protein coding sequence.

DE cytostatic; immunotoxin; cancer; mitochondrial malate dehydrogenase;

XX enzyme; human; MDH, ds; gene; lens crystalline protein.

XX Homo sapiens.

```

XX FH Key Location/Qualifiers
XX FT CDS 1..528
XX FT /*tag= a
XX FT /product= "lens crystalline protein"
XX PN WO2004070012-A2.
XX PD 19-AUG-2004.
XX PF 02-FEB-2004; 2004WO-US002974.
XX PR 02-FEB-2003; 2003US-00444191.
XX PR 03-FEB-2003; 2003US-0444191P.
XX PR 08-APR-2003; 2003US-0460855P.
XX PA (PALO-) PALO ALTO INST MOLECULAR MEDICINE.
XX PI Wright SC, Larrick JW, Nock SR, Wilson DS;
XX KW WPI; 2004-604434/58.
XX DR P-PSDB; ADR46387.
XX PT New compositions comprising proteins and encoding nucleic acids having a
XX PT DNA nuclease or cell killing activity and are operably linked to cancer
XX PT cell binding antibodies or growth factors, useful for treating cancer.
XX PS Disclosure; SEQ ID NO 35; 225pp; English.
XX CC The present invention relates to a composition comprising an isolated
XX CC amino acid sequence that comprises a portion of human mitochondrial
XX CC malate dehydrogenase protein (MDH), particularly the minimum activator of
XX CC DNA fragmentation and activator of DNA fragmentation sequences. The
XX CC composition is useful for treating cancer chosen from liver cancer,
XX CC gastric cancer, head cancer, neck cancer, lung cancer, breast cancer,
XX CC prostate cancer, cervical cancer, pancreatic cancer, colon cancer,
XX CC ovarian cancer, stomach cancer, oesophagus cancer, mouth cancer, tongue
XX CC cancer, gum cancer, skin cancer, muscle cancer, heart cancer, bronchial
XX CC cancer, cartilage cancer, bone cancer, testis cancer, kidney cancer,
XX CC endometrium cancer, uterus cancer, bladder cancer, bone marrow cancer,
XX CC lymphoma cancer, spleen cancer, thymus cancer, thyroid cancer, brain
XX CC cancer, neuron cancer, gall bladder cancer, ocular cancer, joint cancer,
XX CC glioblastoma, mesothelioma, lymphoma, leukaemia, melanoma, squamous cell
XX CC carcinoma, osteosarcoma, and Kaposi's sarcoma. The present sequence is a
XX CC coding sequence shown in the exemplification of the invention.
XX SQ Sequence 528 BP; 115 A; 167 C; 123 G; 123 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 6,166-37 Length: 528
Score: 354.00 Matches: 66
Percent Similarity: 78.45% Conservative: 25
Best Local Similarity: 56.90% Mismatches: 19
Query Match: 55.40% Indels: 6
DB: 13 Gaps: 2

US-10-657-740-1_COPY_51_173 (1-123) x ADR46388 (1-528)

QY 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27
DB 184 GACACTGGACTCTCAGATGGCGCTGGAGAGAGACAGGTTCTCTGTAACCTGGATGTG 243
QY 28 LysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGluIleHis 47
DB 244 AAGCACTTCTCCCGAGAGAACTCAAGTTAAGGTGTGGGAGATGATGATGATGATGAT 303
QY 48 GlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArgArg 67
DB 304 GGAAACATGAGAGCGCCAGCATGACATGTTTCACTCCAGGAGTTCACAGGAA 363
QY 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87
DB 364 TACCGATCCCGATGATGATAGACCTCTCTCACCAATTACTTCTCATCTGTCATGCGG 423

```

88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107  
 424 GTCCTCACTGTGAATGACCAAGGAACAG-----GTCTCTGGCCCTGAGCGC 471  
 108 AlaIleProValSerArgGluLysPro-----ThrSerAlaPro 121  
 472 ACCATTCCCATCACCCTGTGAGAGAGAGCTGTCTGTACCCGAGCCCC 519

RESULT 14  
 ADE75374  
 ID ADE75374 standard; DNA; 537 BP.  
 XX AC ADE75374;  
 XX DT 29-JAN-2004 (first entry)  
 XX DE Human mutant alphaB-crystallin fragment-encoding DNA.  
 XX KW Hybrid protein chaperone; protein stabilisation; heat shock protein;  
 KW SHSP family; protein aggregation inhibition; cell death inhibition;  
 KW genome stability pathway inhibition; protein denaturation identification;  
 KW protein conformation related disease; cardiomyopathy; cataract;  
 KW neurodegenerative disease; cardiac; ophthalmological; neuroprotective;  
 KW gene therapy; alphaB-crystallin; wild-type; human; cassette mutagenesis;  
 KW mutant; ds.  
 XX OS Synthetic.  
 XX OS Homo sapiens.  
 XX FH Key Location/Qualifiers  
 FT misc\_feature /tag= a  
 FT mutation /note= "Unique Aval site for cassette mutagenesis"  
 FT mutation replace(497,T)  
 FT mutation /tag= b  
 FT mutation /note= "Mutagenesis to G from T at this site generates a  
 unique Aval site between bases 493-498"  
 WO2003091266-A2.  
 06-NOV-2003.  
 23-APR-2003; 2003WO-GB001721.  
 23-APR-2002; 2002GB-00009334.  
 (UVDU-) UNIV DUNDEE.  
 Quinlan R;  
 WPI; 2003-865571/80.  
 New hybrid protein chaperone (e.g. heat shock protein) useful for  
 stabilizing proteins and/or protein activities, or as an agent to prevent  
 protein aggregation, or for treating diseases involving altered protein  
 conformations.  
 Disclosure; Fig 12; 45pp; English.  
 The invention relates to a hybrid protein chaperone for stabilising  
 proteins and/or protein activities. Protein chaperones (also known as  
 heat shock proteins) are divided into 4 families on the basis of their  
 primary sequence and chaperone properties: HSP90, HSP70, HSP60 and sHSP  
 (small heat shock protein). The invention is based upon the finding that  
 among the sHSP family, which have a general structure of a central domain  
 (called the alpha-crystallin domain) flanked by N and C-terminal regions,  
 replacement of one or more regions of an sHSP with the corresponding  
 region from a second sHSP can improve the activity compared to native  
 sHSPs. In a particular embodiment of the invention, the hybrid chaperone  
 is a hybrid sHSP designated alphaB-HSP27 comprising the N-terminus and  
 central portion of alphaB-crystallin and the C-terminal tail of HSP27.  
 However, the hybrid protein chaperones of the invention can comprise





```
DB: 10 Gaps: 2
US-10-657-740-1_COPY_51_173 (1-123) x ADE75375 (1-537)
QY 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27
Db 192 GACACTGGACTCTCAGAGATGCGCTGGAGAGGACAGGTTCTCTGTCACCTGGATGG 251
QY 28 LysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGluLeuHis 47
Db 252 AAGCACTTCTCCCGCAGAGAACTCAAGTTAAGGTGTTGGAGATGTGATTGAGGTGCAT 311
QY 48 GlyLysHisAsnGluArgGlnAspAspHisGlyTyrlleSerArgGluPheHisArg 67
Db 312 GGAAACATGAAAGAGCGCCAGATGAACATGTTTCTCATCTCCAGGAGTTCCACAGGAAA 371
QY 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87
Db 372 TACCGATCCAGCTGATGAGCCCTCTCACCATTACTTCATCCCTGTCATCTGATGG 431
QY 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107
Db 432 GTCCTCACTGTGAATGACCAAGGAACAG-----GTCTCTGGCCCTGAGCGC 479
QY 108 AlaIleProValSerArgGluLysPro-----ThrSerAlaPro 121
Db 480 ACCATTCCCATCACCGTGAAGAGAGCGCTGTGTCTCAGCGAGCCCCC 527

RESULT 16
AAX39668
ID AAX39668 standard; DNA; 691 BP.
XX
AC AAX39668;
XX
DT 02-JUL-1999 (first entry)
XX
DE Renal cancer associated gene.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
XX
OS Homo sapiens.
XX
FN WO9904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US014679.
XX
PR 17-JUL-1997; 97US-00896164.
PR 10-OCT-1997; 97US-0061599P.
PR 10-OCT-1997; 97US-0061765P.
PR 10-OCT-1997; 97US-00948705.
PR 11-OCT-1997; 97GB-00021697.
PR 22-JUN-1998; 98US-00102322.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
XX Old LJ, Scanlan MJ, Stockert E, Gure A, Chen Y, Gout I;
PI O'hare M, Obata Y, Pfeundschoh M, Tureci O, Sahin U;
XX
XX WPI; 1999-132448/11.
XX
XX New isolated cancer associated nucleic acids and polypeptides - isolated
PT using sera from cancer patients, used to develop products for the
PT diagnosis, monitoring or treatment of cancers.
XX
PS Claim 67; Page 478; 787pp; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
```

```
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer
XX
```

SQ Sequence 691 BP; 168 A; 208 C; 148 G; 167 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.:	9,07e-37	Length:	691
Score:	354.00	Matches:	66
Percent Similarity:	78.45%	Conservative:	25
Best Local Similarity:	56.90%	Mismatches:	19
Query Match:	55.40%	Indels:	6
DB:	2	Gaps:	2

US-10-657-740-1\_COPY\_51\_173 (1-123) x AAX39668 (1-691)

```
QY 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27
Db 209 GACACTGGACTCTCAGAGATGCGCTGGAGAGGACAGGTTCTCTGTCAACCTGGATGG 268
```

```
QY 28 LysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGluLeuHis 47
Db 269 AAGCACTTCTCCCGCAGAGAACTCAAGTTAAGGTGTTGGAGATGTGATTGAGGTGCAT 328
```

```
QY 48 GlyLysHisAsnGluArgGlnAspAspHisGlyTyrlleSerArgGluPheHisArg 67
Db 329 GGAAACATGAAAGAGCGCCAGGATGAACATGTTTCTCATCTCCAGGAGTTCCACAGGAAA 388
```

```
QY 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87
Db 389 TACCGATCCAGCTGATGAGCCCTCTCACCATTACTTCATCTCCCTGTCTGATGG 448
```

```
QY 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107
Db 449 GTCCTCACTGTGAATGACCAAGGAACAG-----GTCTCTGGCCCTGAGCGC 496
```

QY 108 AlaIleProValSerArgGluLysPro-----ThrSerAlaPro 121

Db 497 ACCATTCCCATCACCGTGAAGAGAGCGCTGTGTCTCAGCGAGCCCCC 544

#### RESULT 17

ABL65204  
ID ABL65204 standard; DNA; 691 BP.

XX ABL65204;

XX 15-MAY-2002 (first entry)

XX Lung cancer related gene sequence SEQ ID NO:3541.

XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;  
KW gene; ds.

XX Homo sapiens.

XX WO200194629-A2.

XX 13-DEC-2001.

XX 30-MAY-2001; 2001WO-US010838.

XX 05-JUN-2000; 2000US-0209473P.

PR 05-JUN-2000; 2000US-0209531P.

PR	18-SEP-2000;	2000US-02331133P.
PR	18-SEP-2000;	2000US-02331133P.
PR	20-SEP-2000;	2000US-0233617P.
PR	20-SEP-2000;	2000US-0234009P.
PR	20-SEP-2000;	2000US-0234034P.
PR	20-SEP-2000;	2000US-0234034P.
PR	22-SEP-2000;	2000US-0234052P.
PR	22-SEP-2000;	2000US-0234509P.
PR	22-SEP-2000;	2000US-0234567P.
PR	25-SEP-2000;	2000US-0234923P.
PR	25-SEP-2000;	2000US-0234924P.
PR	25-SEP-2000;	2000US-0235077P.
PR	25-SEP-2000;	2000US-0235082P.
PR	25-SEP-2000;	2000US-0235113P.
PR	25-SEP-2000;	2000US-0235280P.
PR	26-SEP-2000;	2000US-0235637P.
PR	26-SEP-2000;	2000US-0235638P.
PR	27-SEP-2000;	2000US-0235711P.
PR	27-SEP-2000;	2000US-0235720P.
PR	27-SEP-2000;	2000US-0235840P.
PR	28-SEP-2000;	2000US-0235863P.
PR	28-SEP-2000;	2000US-0236032P.
PR	28-SEP-2000;	2000US-0236033P.
PR	28-SEP-2000;	2000US-0236034P.
PR	28-SEP-2000;	2000US-0236109P.
PR	28-SEP-2000;	2000US-0236111P.
PR	29-SEP-2000;	2000US-0236842P.
PR	29-SEP-2000;	2000US-0236891P.
PR	02-OCT-2000;	2000US-0237172P.
PR	02-OCT-2000;	2000US-0237173P.
PR	02-OCT-2000;	2000US-0237278P.
PR	02-OCT-2000;	2000US-0237294P.
PR	02-OCT-2000;	2000US-0237295P.
PR	03-OCT-2000;	2000US-0237316P.
PR	03-OCT-2000;	2000US-0237425P.
PR	03-OCT-2000;	2000US-0237598P.
PR	03-OCT-2000;	2000US-0237604P.
PR	03-OCT-2000;	2000US-0237606P.
PR	03-OCT-2000;	2000US-0237608P.
PR	01-NOV-2000;	2000US-0244867P.
PR	01-NOV-2000;	2000US-0245084P.
XX	(AVAL-)	AVALON PHARM.
XX	Young PE, Augustus M,	Carter KC, Ebner R, Endress G, Horrigan S;
PI	Soppet DR, Weaver Z;	
XX	WPI; 2002-188264/24.	
XX	Screening for anti-neoplastic agent involves exposing cells to a chemical	
PT	agent to be tested for anti-neoplastic activity, and determining a change	
PT	in expression of a gene of a signature gene set.	
XX	Claim 1; SEQ ID NO 3541; 44pp; English.	
PS	The present invention describes a method (M1) for screening for an anti-	
XX	neoplastic agent. The method involves exposing cells to a chemical agent	
CC	to be tested for anti-neoplastic activity, determining a change in	
CC	expression of at least one gene (I) of a signature gene set, where (I)	
CC	comprises a sequence (S) selected from 8447 sequences (given in ABL61664	
CC	to ABL70110), or is at least 95% identical to (S), where a change in	
CC	expression is indicative of anti-neoplastic activity. (I) has Cytostatic	
CC	activity and can be used in gene therapy. M1 can be used for screening an	
CC	anti-neoplastic agent, and can be used for producing a product which is	
CC	the data collected with respect to the anti-neoplastic agent as a result	
CC	of M1, and the data is sufficient to convey the chemical structure and/or	
CC	properties of the agent. M1 can be used in the treatment of cancer such	
CC	as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,	
CC	prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell	
CC	cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous	
CC	cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms'	
XX	tumour	
XX	Sequence 691 BP; 168 A; 208 C; 148 G; 167 T; 0 U; 0 Other;	

Alignment Scores:									
Pred. No.:	9,076-37	Length:	691						
Score:	354.00	Matches:	66						
Percent Similarity:	78.45%	Conservative:	25						
Best Local Similarity:	56.90%	Mismatches:	19						
Query Match:	55.40%	Indels:	6						
DB:	6	Gaps:	2						
US-10-657-740-1_COPY_51_173 (1-123) x ABL65204 (1-691)									
QY	8	AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal	27						
DB	209	GACACTGGACTCTCAGAGATGGCCCTGGAGAGGACAGGTTCTCTGTCAACCTGGATGG	268						
QY	28	LysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGluIleHis	47						
DB	269	AAGCACTTCTCCCCAGAGGAATCAAAGTTAAGGTGTTGGGAGATCTGATTGAGGTGCAT	328						
QY	48	GlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArg	67						
DB	329	GGAAAAACATGAAGAGCGCCAGATGAACATGGTTTCATCTCCGGAGTTCACAGGAAA	388						
QY	68	TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly	87						
DB	389	TACCGGATCCCGAGCTGATGTAGACCCCTCTCACCATTACTTCATCCCTGTCTCATCTGATGG	448						
QY	88	MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg	107						
DB	449	GTCCTCACTGTGATGGACCAAGGAACAG-----GTCTCTGGCCCTGAGCGC	496						
QY	108	AlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro	121						
DB	497	ACCATTCCCATCACCCGTGAAGAGAAGCGCTGCTGCACCGCAGCCCCC	544						
RESULT 18									
ID	ABL62427								
XX	ABL62427 standard; DNA; 691 BP.								
AC	ABL62427;								
XX									
DT	15-MAY-2002 (first entry)								
XX									
DE	Colon adenocarcinoma related gene sequence SEQ ID NO:764.								
XX									
KW	Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;								
KW	stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;								
KW	cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;								
KW	gene; ds.								
XX									
XX	Homo sapiens.								
XX									
PN	W0200194629-A2.								
XX									
PD	13-DEC-2001.								
XX									
PF	30-MAY-2001; 2001WO-US010838.								
XX									
PR	05-JUN-2000; 2000US-0209473P.								
PR	05-JUN-2000; 2000US-0209531P.								
PR	18-SEP-2000; 2000US-0233133P.								
PR	18-SEP-2000; 2000US-02331617P.								
PR	20-SEP-2000; 2000US-0234009P.								
PR	20-SEP-2000; 2000US-0234034P.								
PR	20-SEP-2000; 2000US-0234052P.								
PR	22-SEP-2000; 2000US-0234509P.								
PR	22-SEP-2000; 2000US-0234567P.								
PR	25-SEP-2000; 2000US-0234923P.								
PR	25-SEP-2000; 2000US-0234924P.								
PR	25-SEP-2000; 2000US-0235077P.								
PR	25-SEP-2000; 2000US-0235082P.								
PR	25-SEP-2000; 2000US-0235134P.								
PR	25-SEP-2000; 2000US-0235280P.								



```
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 691 BP; 168 A; 208 C; 148 G; 167 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 9,07e-37 Length: 691
Score: 354.00 Matches: 66
Percent Similarity: 78.45% Conservative: 25
Best Local Similarity: 56.90% Mismatches: 19
Query Match: 55.40% Indels: 6
DB: 6 Gaps: 2

US-10-657-740-1_COPY_51_173 (1-123) x ABK10966 (1-691)

QY 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27
DB GACACTGGACACTCAGAGATCGCTGGAGAGAGACAGGTTCTCTGTCACCTGGATGTG 268
QY 28 LysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGluIleHis 47
DB AAGCACTTCTCCCGAGAGAACTCAAAAGTTAAGGTGTTGGAGATGTGATGGTGCAT 328
QY 48 GlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArgArg 67
DB GGAAACATGAGAGCCCGCAGGATGAACATGGTTTCATCTCCAGGAGTTCACAGGAAA 388
QY 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87
DB TACCGGATCCAGCTGATGTAGACCTCTCACCATTACTTCTCCTGTCATCTGATGGG 448
QY 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107
DB GTCTCTACTGTGAATGGACCAAGGAACAG-----GTCTCTGGCCCTGAGCGC 496
QY 108 AlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 121
DB ACCATTCCTCATCCCGTGAAGAGAGAGCTGTGTCTACCGAGCCCCC 544

RESULT 20
ID ABK64671 standard; DNA; 691 BP.
XX
AC ABK64671;
DB 18-JUN-2002 (first entry)
XX
DE Human benign prostatic hyperplasia gene #566.
XX
KW Human; benign prostatic hyperplasia; BPH; prostate cancer; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200212440-A2.
XX
PD 14-FEB-2002.
XX
PF 07-AUG-2001; 2001WO-US024708.
XX
PR 07-AUG-2000; 2000US-022323P.
XX
PP 05-JUN-2001; 2001US-00873319.
XX
PA (GENE-) GENE LOGIC INC.
XX
PB (NIBS) JAPAN TOBACCO INC.
XX
PI Munger WE, Kulkarni P, Getzenberg RH, Waga I, Yamamoto J;
XX
PI WPI; 2002-257476/30.
XX
PT Identifying drugs for and diagnosing benign prostatic hyperplasia, by
PT detecting expression levels of one or more genes in prostate cells from
PT patient that are differentially regulated compared to normal prostate
PT cells.
XX
```

```
PS Disclosure; Page 307-308; 444pp; English.
XX
CC The invention relates to a method of diagnosing (I) the onset or
CC progression of benign prostatic hyperplasia (BPH), or screening (II) for
CC or identifying an agent that modulates the onset or progression of BPH.
CC The method is based on changes in gene expression in BPH tissue isolated
CC from patients exhibiting different clinical states of prostate
CC hyperplasia as compared to normal prostate tissue. (I) comprises
CC detecting the expression levels of one or more genes in prostate cells
CC from the subject that are differentially regulated compared to normal
CC prostate cells. (II) comprises preparing a first gene expression profile
CC of BPH cells or BPH-like cell population, exposing the cells to the
CC agent, preparing a second gene expression profile of the agent exposed
CC cells, and comparing the first and second gene expression profiles. (I)
CC is useful for diagnosing the onset or progression of BPH. (II) is useful
CC for identifying an agent that modulates the onset or progression of BPH.
CC The methods are useful to present information identifying the expression
CC level in a tissue or cells, by comparing the expression level of genes
CC given in the specification in the tissue or cells to the level of
CC expression of gene in the database, and displaying the expression levels
CC of at least one gene in the tissue or cell sample compared to the
CC expression level in BPH. Agents using (II) are useful for treating BPH or
CC prostate cancer. ABK64106-ABK64860 represent human benign prostatic
CC hyperplasia gene sequences of the invention
XX
SQ Sequence 691 BP; 168 A; 208 C; 148 G; 167 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 9,07e-37 Length: 691
Score: 354.00 Matches: 66
Percent Similarity: 78.45% Conservative: 25
Best Local Similarity: 56.90% Mismatches: 19
Query Match: 55.40% Indels: 6
DB: 6 Gaps: 2

US-10-657-740-1_COPY_51_173 (1-123) x ABK64671 (1-691)

QY 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27
DB GACACTGGACACTCAGAGATCGCTGGAGAGAGACAGGTTCTCTGTCACCTGGATGTG 268
QY 28 LysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGluIleHis 47
DB AAGCACTTCTCCCGAGAGAACTCAAAAGTTAAGGTGTTGGAGATGTGATGGTGCAT 328
QY 48 GlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArgArg 67
DB GGAAACATGAGAGCCCGCAGGATGAACATGGTTTCATCTCCAGGAGTTCACAGGAAA 388
QY 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87
DB TACCGGATCCAGCTGATGTAGACCTCTCACCATTACTTCTCCTGTCATCTGATGGG 448
QY 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107
DB GTCTCTACTGTGAATGGACCAAGGAACAG-----GTCTCTGGCCCTGAGCGC 496
QY 108 AlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 121
DB ACCATTCCTCATCCCGTGAAGAGAGAGCTGTGTCTACCGAGCCCCC 544

RESULT 21
ID ADF09686 standard; cDNA; 691 BP.
XX
AC ADF09686;
XX
DT 12-FEB-2004 (first entry)
XX
DE Human crystallin, alpha B cDNA SEQ ID NO:187.
XX
KW human; protein-protein interaction; virucide; cytostatic; vaccine;
KW human papilloma virus; HPV; cancer; ss; gene.
```

```
XX OS Homo sapiens.
XX PN WO2003068940-A2.
XX PD 21-AUG-2003.
XX PF 14-FEB-2003; 2003WO-US004594.
XX PR 14-FEB-2002; 2002US-0356911P.
XX PA (CURA-) CURAGEN CORP.
XX PA (HOFF ) HOFFMANN LA ROCHE INC.
XX PI Jackson A, Ooi CE, Lewin DA, Cuthill S;
XX WPI; 2003-689668/65.
XX P-PSDB; ADF09576.
XX
XX New purified complex comprising a first polypeptide and a second
XX polypeptide, useful for identifying agents for treating/preventing a
XX condition involving altered level of the complex e.g. human papilloma
XX virus infection, or cancer.
XX
XX Example 3; SEQ ID NO 187; 156pp; English.
XX
XX The invention relates to a novel purified complex comprising a first
XX polypeptide and a second polypeptide, where the polypeptides comprise
XX defined amino acid sequences listed in the specification, and where the
XX first polypeptide binds to the second polypeptide. A complex of the
XX invention has virucide and cytostatic activity, and may have a use as a
XX vaccine. The complex is useful for identifying agents for treating or
XX preventing a conditions involving altered level of the complex, e.g.
XX human papilloma virus (HPV) infection, or cancer. The compositions,
XX antibodies, vectors and methods are useful for treating such diseases.
XX The sequences shown in ADF09584-ADF09697 represent cDNA's of the
XX invention.
XX
XX SQ Sequence 691 BP; 168 A; 208 C; 148 G; 167 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 9.07e-37 Length: 691
Score: 354.00 Matches: 66
Percent Similarity: 78.45% Conservative: 25
Best Local Similarity: 56.90% Mismatches: 19
Query Match: 55.40% Indels: 6
DB: 10 Gaps: 2

US-10-657-740-1_COPY_51_173 (1-123) x ADF09686 (1-691)
Qy 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27
Db 209 GACACTGGACTCTCAGAGATCGCGCTGGAGAGGACAGGTTCTCTGTCAACCTGGATGTG 268
Qy 28 LysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGluLeHis 47
Db 269 AAGCACTTCTCCCGAGAGAACTCAAAGTTAAGGTGTGGAGATGTGATTGGATGTCAT 328
Qy 48 GlyLysHisAenGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArg 67
Db 329 GGAACAATGATGAAGAGCGCCAGGATGAACATGGTTTCATCTCCAGGGAGTTCACAGGAAA 388
Qy 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87
Db 389 TACCGATCCAGCTGATGATAGACCTCTCACCATTACTTCTCATCTCATCTGATGGG 448
Qy 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107
Db 449 GTCTCTACTGTGAATGACCAAGAAACAG-----GTCTCTGGCCCTGAGCGC 496
Qy 108 AlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 121
Db 497 ACCATTCCCATCACCCCGTGAAGAGAGCGCTCTGTCTCACCAGCCGCC 544
```

```
RESULT 22
ADN05800
ID ADN05800 standard; cDNA; 691 BP.
XX AC ADN05800;
XX AC ADN05800;
XX DT 01-JUL-2004 (first entry)
XX DE Antipsoriatic cDNA sequence #1131.
XX KW ds; gene; antipsoriatic; gene therapy; psoriasis; diagnosis.
XX OS Homo sapiens.
XX PN WO2004028479-A2.
XX PD 08-APR-2004.
XX PF 25-SEP-2003; 2003WO-US030907.
XX PR 25-SEP-2002; 2002US-0414006P.
XX PA (GETH ) GENENTECH INC.
XX PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;
XX PI Wu TD;
XX WPI; 2004-305105/28.
XX P-PSDB; ADN05801.
XX
XX New PRO nucleic acid or polypeptide, useful for preparing a
XX pharmaceutical composition for diagnosing or treating psoriasis in a
XX mammal.
XX
XX Claim 1; SEQ ID NO 2194; 3069pp; English.
XX
XX The invention relates to novel polynucleotide and polypeptides for
XX treating psoriasis or a sequence having at least 80% identity to the
XX above sequences. The nucleic acid is useful for preparing a composition
XX for diagnosing or treating psoriasis in a mammal. This sequence
XX corresponds to one of the polynucleotides of the invention.
XX
XX SQ Sequence 691 BP; 168 A; 208 C; 148 G; 167 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 9.07e-37 Length: 691
Score: 354.00 Matches: 66
Percent Similarity: 78.45% Conservative: 25
Best Local Similarity: 56.90% Mismatches: 19
Query Match: 55.40% Indels: 6
DB: 12 Gaps: 2

US-10-657-740-1_COPY_51_173 (1-123) x ADN05800 (1-691)
Qy 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27
Db 209 GACACTGGACTCTCAGAGATCGCGCTGGAGAGGACAGGTTCTCTGTCAACCTGGATGTG 268
Qy 28 LysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGluLeHis 47
Db 269 AAGCACTTCTCCCGAGAGAACTCAAAGTTAAGGTGTGGAGATGTGATTGGATGTCAT 328
Qy 48 GlyLysHisAenGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArg 67
Db 329 GGAACAATGATGAAGAGCGCCAGGATGAACATGGTTTCATCTCCAGGGAGTTCACAGGAAA 388
Qy 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87
Db 389 TACCGATCCAGCTGATGATAGACCTCTCACCATTACTTCTCATCTCATCTGATGGG 448
Qy 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107
Db 497 ACCATTCCCATCACCCCGTGAAGAGAGCGCTCTGTCTCACCAGCCGCC 544
```



Query Match: 55.40% Indels: 6  
DB: 3 Gaps: 2

US-10-657-740-1\_COPY\_51\_173 (1-123) x AAC03893 (1-856)

Qy 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27  
Db 412 GACACTGGACTCTCAGAGATCGCGCTGGAGAGACAGGTTCTCTGTCAACCTGGATGTG 471

Qy 28 LysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHis 47  
Db 472 AAGCACTTCTCCCGAGAGAACTCAAGTTAAGGTGTGGAGATGTGATGGATGTCAT 531

Qy 48 GlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArg 67  
Db 532 GGAACAACATGAAGAGCGCCAGGATGAACATGGTTTCATCTCCAGGGAGTTCACAGAGAAA 591

Qy 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87  
Db 592 TACCGGATCCAGCTGATGTAGACCTCTCACCATTACTTCTATCCCTGTCTATCTGATGGG 651

Qy 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107  
Db 652 GTCCTCACTGTGAATGGACCAAGAAACAG-----GTCCTCGGCCCTGAGCGC 699

Qy 108 AlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 121  
Db 700 ACCATTCCCATCCCGTGAAGAGAGAGCTCTCTGTCTCAGCGAGCCCC 747

RESULT 25  
AAC10867  
ID AAC10867 standard; cDNA; 893 BP.  
XX AAC10867;  
XX  
XX 06-OCT-2000 (first entry)  
XX  
XX Human secreted protein 5' EST, SEQ ID NO: 14942.  
XX  
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
XX gene therapy; chromosome mapping; ss.  
XX  
XX Homo sapiens.  
XX  
XX EP1033401-A2.  
XX  
XX 06-SEP-2000.  
XX  
XX 21-FEB-2000; 2000EP-00200610.  
XX  
XX 26-FEB-1999; 99US-0122487P.  
XX  
XX (GENSET).  
XX  
XX Dumas Milne Edwards J, Duclert A, Giordano J;  
XX  
XX WPI; 2000-500381/45.  
XX  
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
XX diagnostic, forensic, gene therapy and chromosome mapping procedures.  
XX  
XX Claim 1; SEQ ID NO 14942; 71pp + Sequence Listing; English.  
XX  
XX The present sequence is one of a large number of 5' ESTs derived from  
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively  
XX identified within the present sequence. The 5' ESTs were prepared from  
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
XX sequences usually correspond mainly to the 3' untranslated region (UTR)  
XX of the mRNA because they are often obtained from oligo-dT primed cDNA  
XX libraries. Such ESTs are not well suited for isolating cDNA sequences  
XX derived from the 5' ends of mRNAs and even in those cases where longer  
XX cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'

CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used  
CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in  
CC diagnostic, forensic, gene therapy and chromosome mapping procedures.  
CC They are used to obtain upstream regulatory sequences and to design  
CC expression and secretion vectors  
XX  
XX Sequence 893 BP; 221 A; 259 C; 215 G; 198 T; 0 U; 0 Other;  
SQ

Alignment Scores:  
Pred. No.: 1.31e-36 Length: 893  
Score: 354.00 Matches: 66  
Percent Similarity: 78.45% Conservative: 25  
Best Local Similarity: 56.90% Mismatches: 19  
Query Match: 55.40%  
DB: 3 Gaps: 2

US-10-657-740-1\_COPY\_51\_173 (1-123) x AAC10867 (1-893)

Qy 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27  
Db 449 GACACTGGACTCTCAGAGATCGCGCTGGAGAGACAGGTTCTCTGTCAACCTGGATGTG 508

Qy 28 LysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHis 47  
Db 509 AAGCACTTCTCCCGAGAGAACTCAAGTTAAGGTGTGGAGATGTGATGGATGTCAT 568

Qy 48 GlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArg 67  
Db 569 GGAACAACATGAAGAGCGCCAGGATGAACATGGTTTCATCTCCAGGGAGTTCACAGAGAAA 628

Qy 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87  
Db 629 TACCGGATCCAGCTGATGTAGACCTCTCACCATTACTTCTATCCCTGTCTATCTGATGGG 688

Qy 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107  
Db 689 GTCCTCACTGTGAATGGACCAAGAAACAG-----GTCCTCGGCCCTGAGCGC 736

Qy 108 AlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 121  
Db 737 ACCATTCCCATCCCGTGAAGAGAGAGCTCTGTCTCAGCGAGCCCC 784

RESULT 26  
AAC10863  
ID AAC10863 standard; cDNA; 911 BP.  
XX AAC10863;  
XX  
XX 06-OCT-2000 (first entry)  
XX  
XX Human secreted protein 5' EST, SEQ ID NO: 14938.  
XX  
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
XX gene therapy; chromosome mapping; ss.  
XX  
XX Homo sapiens.  
XX  
XX OS  
XX  
XX EP1033401-A2.  
XX  
XX 06-SEP-2000.  
XX  
XX 21-FEB-2000; 2000EP-00200610.  
XX  
XX 26-FEB-1999; 99US-0122487P.  
XX  
XX (GENSET).  
XX  
XX Dumas Milne Edwards J, Duclert A, Giordano J;  
XX  
XX WPI; 2000-500381/45.  
XX  
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for

PT diagnostic, forensic, gene therapy and chromosome mapping procedures.  
XX Claim 1; SEQ ID NO 14938; 71pp + Sequence Listing; English.  
XX  
XX The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
CC identified within the present sequence. The 5' ESTs were prepared from  
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
CC derived from the 5' ends of mRNAs and even in those cases where longer  
CC cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'  
CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used  
CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in  
CC diagnostic, forensic, gene therapy and chromosome mapping procedures.  
CC They are used to obtain upstream regulatory sequences and to design  
CC expression and secretion vectors  
XX  
SQ Sequence 911 BP; 221 A; 260 C; 211 G; 219 T; 0 U; 0 Other;  
  
Alignment Scores:  
Pred. No.: 1.35e-36 Length: 911  
Score: 354.00 Matches: 66  
Percent Similarity: 78.45% Conservative: 25  
Best Local Similarity: 56.90% Mismatches: 19  
Query Match: 55.40% Indels: 6  
DB: 3 Gaps: 2  
  
US-10-657-740-1\_COPY\_51\_173 (1-123) x AAC10863 (1-911)  
  
QY 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27  
DB 467 GACACTGGACTCTCAGAGATGGCTGGAGAGACAGGTTCTGTCAACTGGATGTG 526  
  
QY 28 LysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHis 47  
DB 527 AAGCACTTCTCCCGACAGAACTCAAGTAAAGTGTGGGAGATGTGATTGAGTGCAT 586  
  
QY 48 GlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArg 67  
DB 587 GGAACAACTAAGAGCGCCAGATGAACATGGTTTCATCTCCAGGGAGTTCACAGGAAA 646  
  
QY 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87  
DB 647 TACCGATCCAGCTGATGTAGACCTCTCACCATTACTTCTATCCCTGCTCATCTGATGG 706  
  
QY 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107  
DB 707 GTCTCTCACTGTGAATGGACCAAGGAACAG-----GTCTCTGGCCCTGAGGCG 754  
  
QY 108 AlaIleProValSerArgGluLysPro-----ThrSerAlaPro 121  
DB 755 ACCATTCCTCATCCCGTGAAGAGAGCCCTGTGTCCCGAGCCCCC 802  
  
RESULT 27  
AAC10862  
ID AAC10862 standard; cDNA; 913 BP.  
XX  
XX AAC10862;  
AC  
XX  
XX 06-OCT-2000 (first entry)  
DT  
XX  
XX Human secreted protein 5' EST, SEQ ID NO: 14937.  
DE  
XX  
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX EP1033401-A2.  
PN  
XX  
XX 06-SEP-2000.  
PD

XX 21-FEB-2000; 2000EP-00200610.  
PF  
XX 26-FEB-1999; 99US-0122487P.  
PR  
XX (GEST ) GENSET.  
XX  
XX Dumas Milne Edwards J, Duclert A, Giordano J;  
XX  
XX WPI; 2000-500381/45.  
DR  
XX  
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.  
XX  
XX Claim 1; SEQ ID NO 14937; 71pp + Sequence Listing; English.  
XX  
XX The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
CC identified within the present sequence. The 5' ESTs were prepared from  
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
CC derived from the 5' ends of mRNAs and even in those cases where longer  
CC cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'  
CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used  
CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in  
CC diagnostic, forensic, gene therapy and chromosome mapping procedures.  
CC They are used to obtain upstream regulatory sequences and to design  
CC expression and secretion vectors  
XX  
SQ Sequence 913 BP; 228 A; 257 C; 220 G; 202 T; 0 U; 6 Other;  
  
Alignment Scores:  
Pred. No.: 1.35e-36 Length: 913  
Score: 354.00 Matches: 66  
Percent Similarity: 78.45% Conservative: 25  
Best Local Similarity: 56.90% Mismatches: 19  
Query Match: 55.40% Indels: 6  
DB: 3 Gaps: 2  
  
US-10-657-740-1\_COPY\_51\_173 (1-123) x AAC10862 (1-913)  
  
QY 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27  
DB 469 GACACTGGACTCTCAGAGATGGCTGGAGAGACAGGTTCTGTCAACTGGATGTG 528  
  
QY 28 LysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHis 47  
DB 529 AAGCACTTCTCCCGACAGAACTCAAGTAAAGTGTGGGAGATGTGATTGAGTGCAT 588  
  
QY 48 GlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArg 67  
DB 589 GGAACAACTAAGAGCGCCAGATGAACATGGTTTCATCTCCAGGGAGTTCACAGGAAA 648  
  
QY 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87  
DB 649 TACCGATCCAGCTGATGTAGACCTCTCACCATTACTTCTATCCCTGCTCATCTGATGG 708  
  
QY 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107  
DB 709 GTCTCTCACTGTGAATGGACCAAGGAACAG-----GTCTCTGGCCCTGAGGCG 756  
  
QY 108 AlaIleProValSerArgGluLysPro-----ThrSerAlaPro 121  
DB 757 ACCATTCCTCATCCCGTGAAGAGAGCCCTGTGTCCCGAGCCCCC 804  
  
RESULT 28  
AAC10861  
ID AAC10861 standard; cDNA; 927 BP.  
XX  
XX AAC10861;  
AC





Db 498 GACACTGCTCTCAGAGATCGCGCTGGAGAAGCACAGGTTCTCTGTCAACCTGGATGTG 557  
Qy 28 LysHisPheSerProGluAspLeuThrValIysValGlnAspAspPheValGluIleHis 47  
Db 558 AAGACACTTCTCCCGAGAGAACTCAAGTTAAGGTGTGGAGATGTGATTGAGGTGCAT 617  
Qy 48 GlyIysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArg 67  
Db 618 GGAACACATGAAGAGCCAGGATGAACATGTTTCATCTCCAGGAGTTCCACAGAAA 677  
Qy 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87  
Db 678 TACCGATCCAGCTGATGTAGACCCCTCTCACCATTTCTATCCCTGTCTCATCTGATGG 737  
Qy 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107  
Db 738 GTCTCTACTGTGAATGACCAAGAAACAG-----GTCTCTGGCCCTGAGCGC 785  
Qy 108 AlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 121  
Db 786 ACCATTCCCATCACCGTGAAGAGAAGCTGTCTGTCTCACCGCAGCCCCC 833  
RESULT 30  
AAZ33574/C  
ID AAZ33574 standard; cDNA; 1036 BP.  
XX  
AC AAZ33574;  
XX  
DT 08-DEC-1999 (first entry)  
XX  
DE Human breast tumour-associated EST 34.  
XX  
KW Expressed sequence tag; EST; human; breast; cancer; cytostatic;  
KW medicaments; gene therapy; treatment; fat metabolism; ss.  
XX  
OS Homo sapiens.  
XX  
PN DE19013835-AL.  
XX  
PD 23-SEP-1999.  
XX  
PF 20-MAR-1998; 98DE-01013835.  
XX  
PR 20-MAR-1998; 98DE-01013835.  
XX  
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.  
XX  
PI Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;  
XX  
DR WPI; 1999-528979/45.  
XX  
PT Human nucleic acid sequences and protein products from normal breast  
PT tissue, useful for breast cancer therapy.  
XX  
PS Claim 3; 122; 206pp; German.  
XX  
CC This invention describes novel human nucleic acid sequences from normal  
CC breast tissue which have cytostatic activity. The nucleic acid sequences  
CC can be used to produce and isolate full-length gene sequences. They can  
CC be used to express proteins, which can be used as tools to find an  
CC activity against breast cancer. The sequences can be used in sense or  
CC antisense form. They are especially useful for medicaments for gene  
CC therapy to treat breast cancer and for treating illnesses associated with  
CC fat metabolism. AAZ33541-233610 represent expressed sequence tags  
CC described in the method of the invention  
XX  
SQ Sequence 1036 BP; 240 A; 251 C; 283 G; 262 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 1.62e-36 Length: 1036  
Score: 354.00 Matches: 66  
Percent Similarity: 78.45% Conservative: 25  
Best Local Similarity: 56.90% Mismatches: 19

Query Match: 55.40% Indels: 6  
DB: 2 Gaps: 2  
US-10-657-740-1\_COPY\_51\_173 (1-123) x AAZ33574 (1-1036)  
Qy 8 AspSerGlyIleSerGluValArgSerAspArgAspIysPheValIlePheLeuAspVal 27  
Db 524 GACACTGACTCTCAGAGATCGCGCTGGAGAAGCACAGGTTCTCTGTCAACCTGGATGTG 465  
Qy 28 LysHisPheSerProGluAspLeuThrValIysValGlnAspAspPheValGluIleHis 47  
Db 464 AAGCACTTCTCCCGAGAGAACTCAAGTTAAGGTGTGGAGATGTGATTGAGGTGCAT 405  
Qy 48 GlyIysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArg 67  
Db 404 GGAACACATGAAGAGCCAGGATGAACATGTTTCATCTCCAGGAGTTCCACAGAAA 345  
Qy 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87  
Db 344 TACCGATCCAGCTGATGTAGACCCCTCTCACCATTTCTATCCCTGTCTCATCTGATGG 285  
Qy 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107  
Db 284 GTCTCTACTGTGAATGACCAAGAAACAG-----GTCTCTGGCCCTGAGCGC 237  
Qy 108 AlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 121  
Db 236 ACCATTCCCATCACCGTGAAGAGAAGCTGTCTGTCTCACCGCAGCCCCC 189  
RESULT 31  
ABX38978  
ID ABX38978 standard; cDNA; 380 BP.  
XX  
AC ABX38978;  
XX  
DT 20-FEB-2003 (first entry)  
XX  
DE Bovine EST associated with lactation/muscle/fat deposition #4143.  
XX  
KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;  
KW muscle deposition; fat deposition; genome mapping; gene identification;  
KW gene analysis; cattle breeding.  
XX  
OS Bos Taurus.  
XX  
PN US2002137139-A1.  
XX  
PD 26-SEP-2002.  
XX  
PF 24-SEP-2001; 2001US-00960352.  
XX  
PR 12-JAN-1999; 99US-0115707P.  
PR 11-JAN-2000; 2000US-00480902.  
XX  
PA (BYAT/) BYATT J C.  
PA (MATH/) MATHIALAGAN N.  
PA (TAON/) TAO N.  
PA (WARR/) WARREN W C.  
XX  
PI Byatt JC, Mathialagan N, Tao N, Warren WC;  
XX  
DR WPI; 2003-110599/10.  
XX  
PT New nucleic acid associated with lactation, and muscle and fat  
PT deposition, useful for genome mapping, gene identification and analysis,  
PT cattle breeding, or for genetically improving cattle.  
XX  
PS Claim 2; SEQ ID NO 4143; 245pp; English.  
XX  
CC The invention relates to a purified nucleic acid molecule associated with  
CC lactation or muscle and fat deposition (designated LMFD), derived from  
CC cattle, and the LMFD nucleic acid can specifically hybridise to a second  
CC nucleic acid molecule comprising any of 15112 nucleotide sequences,

CC appearing as ABX34836-ABX49947, or complements of them. Also included are  
CC ; (1) a transformed cell having a nucleic acid comprising an LMPD nucleic  
CC acid linked to a promoter and a 3' non-translated sequence that  
CC functions in the cell to cause termination of transcription and addition  
CC of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and  
CC (2) determining a level or pattern of a molecule in a bovine cell or  
CC tissue comprising: (a) incubating a marker nucleic acid (comprising any  
CC of the 15112 nucleic acid sequences or its complement or fragment) with a  
CC complementary nucleic acid molecule obtained from the bovine cell or  
CC tissue, where hybridisation between the marker nucleic acid and the  
CC complementary nucleic acid permits the detection of the molecule; and (b)  
CC detecting the level or pattern of the complementary nucleic acid; where  
CC the detection of the complementary nucleic acid is predictive of the  
CC level or pattern of the molecule. The LMPD nucleic acid is used for  
CC determining a level or pattern of a molecule in a bovine cell or tissue.  
CC It is useful for genome mapping, gene identification and analysis, cattle  
CC breeding, preparation of constructs for use in cattle gene expression, or  
CC for genetically improving cattle. The present sequence is one of the  
CC 15112 bovine LMPD EST (expressed sequence tag) nucleic acids. Note: The  
CC present sequence was not shown in the specification but was obtained in  
CC electronic format from the USPTO web site:  
CC seqdata.uspto.gov/sequence.html?DocID=20020137139  
XX  
SQ Sequence 380 BP; 90 A; 114 C; 101 G; 75 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.:	2.66e-34	Length:	380
Score:	332.50	Matches:	65
Percent Similarity:	75.65%	Conservative:	22
Best Local Similarity:	56.52%	Mismatches:	23
Query Match:	52.03%	Indels:	5
DB:	8	Gaps:	2

US-10-657-740-1\_COPY\_51\_173 (1-123) x ABX38978 (1-380)

Qy	4	ArgThrValLeuAsp	---SerGlyIleSerGluValArgSerAspArgAspLysPheVal	22
Db	12	CGCACCCAGCTGGATGCTCTGCCCTCAGAGATGCGCTGGAGAGAGAGATCTCT	71	
Qy	23	IlePheLeuAspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAsp	42	
Db	72	GTCAACCTGGACCTGACGCACTCTCCACAGAGACTCAAGGCCAAGTCTGGAGAT	131	
Qy	43	PheValGluIleHisGlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArg	62	
Db	132	GTGATTGAGTGCATGCGCAACATCAAGAGCGCCAGGATGAACATGGTTTATCTCCCGG	191	
Qy	63	GluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSer	82	
Db	192	GAGTTCACAGAAATACCGATCCAGCTGACGTGGACCTCTCGCCATTACTTCATCC	251	
Qy	83	LeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAla	102	
Db	252	CTGTCTCTGATGGGACCTACTGTGATGAGCCAGCAAGAACAG-----GCC	299	
Qy	103	ThrHisAlaGluArgAlaIleProValSerArgGluGluLysPro	117	
Db	300	TCCGCCCTTGAGCGCACCATTCCTCCATACCCGTGAAGAGAGCGC	344	

## RESULT 32

AAC03894

ID AAC03894 standard; cDNA; 695 BP.

XX

AC AAC03894;

XX

DT 06-OCT-2000 (first entry)

XX

DE Human secreted protein 5' EST, SEQ ID NO: 3892.

XX

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

XX

KW gene therapy; chromosome mapping; ss.

XX

OS Homo sapiens.

XX EPI033401-A2.  
XX 06-SEP-2000.  
XX 21-FEB-2000; 2000EP-00200610.  
XX 26-FEB-1999; 99US-0122487P.  
XX (GBST ) GENSET.  
XX Dumas Milne Edwards J, Duclert A, Giordano J;  
XX WPI; 2000-500381/45.  
XX P-PSDB; AAG03888.  
XX  
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
XX obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
XX diagnostic, forensic, gene therapy and chromosome mapping procedures.  
XX  
XX Claim 1; SEQ ID NO 3892; 71pp + Sequence Listing; English.  
XX  
XX The present sequence is one of a large number of 5' ESTs derived from  
XX mRNAs encoding secreted proteins. An ORF has been identified within the  
XX sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs  
XX derived from 30 different tissues. EST sequences usually correspond  
XX mainly to the 3' untranslated region (UTR) of the mRNA because they are  
XX often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
XX well suited for isolating cDNA sequences derived from the 5' ends of  
XX mRNAs and even in those cases where longer cDNA sequences have been  
XX obtained the full 5' UTR is rarely included. 5' ESTs are derived from  
XX mRNAs with intact 5' ends and can therefore be used to obtain full length  
XX cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
XX gene therapy and chromosome mapping procedures. They are used to obtain  
XX upstream regulatory sequences and to design expression and secretion  
XX vectors  
XX  
XX Sequence 695 BP; 183 A; 165 C; 162 G; 183 T; 0 U; 2 Other;

## Alignment Scores:

Pred. No.:	1.36e-33	Length:	695
Score:	330.00	Matches:	62
Percent Similarity:	77.27%	Conservative:	23
Best Local Similarity:	56.36%	Mismatches:	19
Query Match:	51.64%	Indels:	6
DB:	3	Gaps:	2

US-10-657-740-1\_COPY\_51\_173 (1-123) x AAC03894 (1-695)

Qy	14	ValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPheSerProGlu	33
Db	269	ATGGCGCTGGAGAGGACAGGTTCTCTCAACCTGGATGTGAAGCACTTCTCCCGAG	328
Qy	34	AspLeuThrValLysValGlnAspPheValGluIleHisGlyLysHisAsnGluArg	53
Db	329	GAACCTCAAGTTAAGGTGTGGGAGATGTGATCAGGTGTCATGGAACATCAAGAGCGC	388
Qy	54	GlnAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeuProSerAsn	73
Db	389	CAGGATGACATGTTTCACTCCAGGAGTTCACAGAGAAATACCGATCCAGCTGAT	448
Qy	74	ValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGly	93
Db	449	GTAGACCTCTCACCATTACTTCTCTGTCATCTGATGGGTCTCTCCTGTGAATGGA	508
Qy	94	ProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProValSerArg	113
Db	509	CCAAAGGAACACAG-----GTCTCGCCCTGAGCGCCACCATTCCTCCACCGT	556
Qy	114	GluGluLysPro-----ThrSerAlaPro	121
Db	557	GAAGAGAGAGCTGTGTCTACCGCGACGCCCC	586



CC (phenotypic probe) R3, Perturbagen F802, Perturbagen F820, their  
 CC biologically active modifications, or biologically active fragments. Also  
 CC including the polynucleotides encoding the perturbagens, a vector  
 CC comprising the polynucleotide, preparing an RA pathway related  
 CC polypeptide, a composition comprising the polypeptide, an antibody to the  
 CC polypeptide, screening (M) putative RA-related therapeutics, by exposing  
 CC a polypeptide/target interaction pair identified by a method utilising  
 CC the RA pathway polypeptide, to a number of agents, and recovering a  
 CC subpopulation of disrupting agents which competitively displace the  
 CC polypeptide from the target, where the disrupting agents are putative RA-  
 CC related therapeutics. Also include are an isolated RA pathway polypeptide  
 CC comprising PARI (a kinesin light chain-related protein) polypeptide and  
 CC its encoding polynucleotide, a gene therapy vector comprising the RA  
 CC pathway protein polypeptide or encoding or PARI and a host cell  
 CC comprising the gene therapy vector. The RA pathway polypeptide is useful  
 CC for identifying a cellular target that interacts with RA pathway-related  
 CC polypeptide, by exposing the polypeptide in vitro to putative target  
 CC molecules and identifying a polypeptide/target interaction pair, by  
 CC detecting reporter expression, where the reporter expression is  
 CC operatively linked to the formation of the interaction pair. (M) is a  
 CC yeast two-hybrid assay. The polypeptide is also useful for treating an RA  
 CC pathway-related condition e.g. lung cancer, Kaposi's sarcoma, breast  
 CC cancer, pancreatic cancer, neuroblastoma, renal cancer, ovarian cancer,  
 CC dermatitis, hyperkeratosis, eczema, Darier's disease, Reiter's disease,  
 CC psoriasis, acute promyelocytic leukaemia (APL). The present sequence is a  
 CC cDNA fragment encoding a partial target molecule isolated by the yeast  
 CC two hybrid system, alpha crystallin B chain

SQ Sequence 349 BP; 75 A; 98 C; 87 G; 89 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 349 Length: 349  
 Score: 263.00 Matches: 44  
 Percent Similarity: 88.24% Conservative: 16  
 Best Local Similarity: 64.71% Mismatches: 8  
 Query Match: 41.16% Indels: 0  
 DB: 6 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x ABK87386 (1-349)

QY 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27  
 DB 132 GACATGGACTCTCAGAGATGCGCTGGAGAGGACAGGTTCTCTGTCACCTGGATGG 191  
 QY 28 LysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGluIleHis 47  
 DB 192 AAGCACTTCTCCCGAGAGAACTCAAGTTAAGTGTGGAGATGTGATGAGTGCAT 251  
 QY 48 GlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArg 67  
 DB 252 GGAAACACATGAAGAGCGCCAGATGAACATGTTTCATCTCCAGGAGTTCACAGGAAA 311  
 QY 68 TyrArgLeuProSerAsnValAsp 75  
 DB 312 TACCGATCCAGCTGATGATAGAC 335

RESULT 35  
 ADM31647

ID ADM31647 standard; cDNA; 615 BP.

AC ADM31647;

DT 17-JUN-2004 (first entry)

DE Human Hsp86 (heat shock protein 27) cDNA.

XX anticancer; head; neck cancer; cytostatic; cisplatin-resistant;

KW antisense therapy; human; ss; Hsp27; heat shock protein 27.

XX Homo sapiens.

XX KR2003065206-A.

XX

PD 06-AUG-2003.

XX 31-JAN-2002; 2002KR-00005714.

XX 31-JAN-2002; 2002KR-00005714.

XX (DNAD-) DNA INC.

XX Jung DG, Kim SB, Lee DS, Park JU;

XX WPI; 2004-117343/12.

XX Antisense oligonucleotide useful in the treatment of cisplatin-resistant  
 XX head and neck cancer.

XX Claim 4; SEQ ID NO 2; 18pp; Korean.

XX The invention relates to a novel anticancer activity-increasing agent for  
 XX enhancing a therapeutic effect on head and neck cancer by inhibiting a  
 XX gene whose expression is increased in head and neck cancer cells. The  
 XX agent of the invention demonstrates cytostatic activities and may be  
 XX useful for enhancing a therapeutic effect on head and neck cancer by  
 XX inhibiting a gene whose expression is increased in head and neck cancer  
 XX cells, as well as for effectively treating cisplatin-resistant head and  
 XX neck cancer, possibly via antisense therapy. The current sequence is that  
 XX of the human Hsp27 (heat shock protein 27) cDNA of the invention.

SQ Sequence 615 BP; 105 A; 234 C; 188 G; 88 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 1,288-24 Length: 615  
 Score: 261.50 Matches: 55  
 Percent Similarity: 65.18% Conservative: 18  
 Best Local Similarity: 49.11% Mismatches: 30  
 Query Match: 40.92% Indels: 10  
 DB: 12 Gaps: 1

US-10-657-740-1\_COPY\_51\_173 (1-123) x ADN31647 (1-615)

QY 7 LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26  
 DB 241 CTCACGCGGGTCTCGGAGATCGGCACACTGCGGACCGTGGCGCGTGTCCCTGGAT 300  
 QY 27 ValLysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGluIle 46  
 DB 301 GTCAACCACTTCGCCCGGACGAGCTGACGGTCAAGACCCAGAGTGGCGTGGAGATC 360  
 QY 47 HisGlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArg 66  
 DB 361 ACCGCAACGACGAGGAGCGGCGGACGAGCATGCTACATCTCCCGGTCTTCACGCGG 420  
 QY 67 ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86  
 DB 421 AAATACACGTGCCCCCGGTGTGGACCCACCAAGTTCTCTCTCTGTCCCTGTGAG 480  
 QY 87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106  
 DB 481 GGCACACTGACCGTGA-GGCCCC-----CATGCCCAA 512  
 QY 107 ArgAlaIleProValSerArgGluLysProThr 118  
 DB 513 GCTAGCCAGCAGTCCACAGAGATCATCCATCCCACT 548

RESULT 36

ADM94741

ID ADM94741 standard; cDNA; 764 BP.

XX ADM94741;

XX 01-JUL-2004 (first entry)

XX Human heat shock protein 27 (hsp27) cDNA sequence SEQ ID NO:91.

XX

```
KW heat shock protein 27; hsp27; cytoskeletal; gene therapy;
KW heat shock protein 27 inhibitor; hsp27 inhibitor; cancer; human; gene;
XX ss.
XX Homo sapiens.
XX WO2004030660-A2.
XX 15-APR-2004.
XX 02-OCT-2003; 2003WO-CA001588.
XX 02-OCT-2002; 2002US-0415859P.
XX 18-APR-2003; 2003US-0463952P.
XX (UYBR-) UNIV BRITISH COLUMBIA.
XX PA Gleave ME, Rocchi P, Signaevsky M;
XX WIPI; 2004-316331/29.
XX DR New composition comprising a therapeutic agent that reduces the amount of
XX PT active hsp27 in hsp27 expressing cells exposed to the therapeutic agent,
XX PT useful in treating cancer, e.g., prostate cancer or a central nervous
XX PT system malignancy.
XX PS Disclosure; SEQ ID NO 91; 38pp; English.
XX CC The present invention describes a composition which comprises a
XX CC therapeutic agent that reduces the amount of active heat shock protein 27
XX CC (hsp27) in hsp27 expressing cells exposed to the therapeutic agent. The
XX CC composition has cytostatic activity, and can be used in gene therapy. The
XX CC composition is useful in treating cancer, e.g., prostate, bladder, lung,
XX CC breast, pancreatic, colon, skin (for example melanoma), renal or ovarian
XX CC cancer or a central nervous system malignancy. The present sequence
XX CC represents a cDNA sequence of human hsp27, which is used in the
XX CC exemplification of the present invention.
XX SQ Sequence 764 BP; 136 A; 283 C; 219 G; 126 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.76e-24 Length: 764
Score: 261.50 Matches: 55
Percent Similarity: 65.18% Conservative: 18
Best Local Similarity: 49.11% Mismatches: 30
Query Match: 40.92% Indels: 10
DB: 12 Gaps: 1

US-10-657-740-1_COPY_51_173 (1-123) x ADM94741 (1-764)
QY 7 LeuAspSerGlyTleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26
Db 266 CTCAGCAGCGGGTCTCGGAGATCCGGACACTGGCGACCGGTGGCGTCTGCTGGAT 325
QY 27 ValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValIle 46
Db 326 GTCAACCACTTCCCGCGAGCGTGGCTGCTCAAGACACAGAGTGGCGTGGAGATC 385
QY 47 HisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArg 66
Db 386 ACCGGCAGCAGCAGGAGCGGCGAGCAGCATGCTCATCTCCGGTGGCTTCCACGGG 445
QY 67 ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86
Db 446 AATACACCGTCCCGCGGTGGAGCCCGACCCCAAGTTCTCTCCCTGCTCCCTCGAG 505
QY 87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106
Db 506 GGCACACTGACCGTGA-GGCCCC-----CATGCCCAA 537
QY 107 ArgAlaIleProValSerArgGluGluLysProThr 118
Db 538 GTACGCCAGCAGTCCACGAGATCACCATCCAGT 573
```

## RESULT 37

```
ADI31974
ID ADI31974 standard; cDNA; 789 BP.
XX AC ADI31974;
XX DT 17-JUN-2004 (first entry)
XX DE Human cDNA #1300.
XX KW Human; gene; ss; immunological response; immunopathological condition;
KW Crohn's disease; asthma; ulcerative colitis; hyper eosinophilia;
KW Irritable bowel syndrome; osteoarthritis; rheumatoid arthritis;
KW acute monocytic leukaemia; antiinflammatory; antirheumatic; antiulcer;
KW osteopathic; antiarthritic; antirheumatic; cytostatic.
XX OS Homo sapiens.
XX PN US6607879-B1.
XX PD 19-AUG-2003.
XX PF 09-FEB-1998; 98US-00023655.
XX PR 09-FEB-1998; 98US-00023655.
XX (INCY-) INCYTE CORP.
XX Cocks BG, Stuart SG, Seilhamer JJ;
XX WIPI; 2003-895307/82.
XX A composition comprising a plurality of cDNAs, useful for detecting
XX altered expression of genes in an immunological response or for
XX diagnosing and treating an immunopathology, e.g. Crohn's disease, asthma
XX or osteoarthritis.
XX Claim 1; SEQ ID NO 1300; 50pp; English.

The invention relates to a composition comprising a plurality of cDNAs
for detecting the altered expression of genes in an immunological
response. The invention also relates to a method of diagnosing or
monitoring the treatment of an immunopathological condition in a sample,
comprising obtaining nucleic acids from a sample, contacting the nucleic
acids of the sample with an array comprising the plurality of cDNAs under
conditions to form one or more hybridisation complexes, detecting the
hybridisation complexes and comparing the levels of the detected
hybridisation complexes with the level of hybridisation complexes
detected in a non-diseased sample, where an altered level of the detected
hybridisation complexes correlates with the presence of an
immunopathological condition. Also disclosed are an expression profile
comprising a microarray and a plurality of detectable complexes and a
method for identifying a plurality of polynucleotide probes. The cDNAs
are useful as hybridisable array elements in a microarray for monitoring
the expression of target polynucleotides. The microarray can be used in
the diagnosis of an immunopathology, such as Crohn's disease, asthma,
ulcerative colitis, hyper eosinophilia, irritable bowel syndrome,
osteoarthritis, rheumatoid arthritis or acute monocytic leukaemia, and in
identifying agents for the treatment of the diseases. The microarray may
also be used in drug discovery and development, toxicological and
carcinogenicity studies, forensics or pharmacogenomics. The composition
may also be used in purification of a subpopulation of mRNAs, cDNAs or
genomic fragments. This sequence represents a human cDNA of the
invention. Note: The sequence data for this patent did not form part of
the printed specification but was obtained in electronic format directly
from USPIO at seqdata.uspto.gov/sequence.html.
XX SQ Sequence 789 BP; 140 A; 296 C; 222 G; 131 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.84e-24 Length: 789
Score: 261.50 Matches: 55
```

Percent Similarity:	65.18%	Conservative:	18
Best Local Similarity:	49.11%	Mismatches:	30
Query Match:	40.92%	Indels:	10
DB:	11	Gaps:	1
US-10-657-740-1_COPY_51_173 (1-123) x ADI31974 (1-789)			
Qy	7	LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp	266
Db	293	CTCAGCAGCGGGTCTCGGAGATCCGGCACACTCGGCACCGGTGGCGCGTGTCCTTGAT	352
Qy	27	VallysHisPheSerProGluAspLeuThrVallysValGlnAspAspPheValGluIle	46
Db	353	GTCAACCACTTCGCCGCCGACGAGCTGACGGTCAAGACCAGGATCGCGTGTTGGAGATC	412
Qy	47	HisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrlIleSerArgGluPheHisArg	66
Db	413	ACCGCAACGACGAGGAGCGGCAGACGACGATGCTCATCTCCCGGTGCTTCACGGCG	472
Qy	67	ArgTyArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp	86
Db	473	AAATACAGCTGCCCGCGGTGTGAACCCCAAGTTTCTCTCCCTGHCCTCCCTGAG	532
Qy	87	GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGlu	106
Db	533	GGCACACTGACCGTGA-GGCCCC-----CATGCCCAA	564
Qy	107	ArgAlaIleProValSerArgGluGluLysProThr	118
Db	565	GCTAGCCACGCGAGTCCAACGAGATCACCATTCCAGT	600

RESULT 38	
ADG10693	
ID	ADG10693 standard; cDNA; 847 BP.
XX	
AC	ADG10693;
XX	
XX	
DT	26-FEB-2004 (first entry)
XX	
DE	Human STAT6-activating protein-encoding cDNA, SEQ ID NO:283.
XX	
KW	Human; STAT6-activating protein; drug screening; activator; inhibitor;
KW	allergic disease; inflammation; autoimmune disease; diabetes;
KW	hyperlipidaemia; cancer; infection; HIV infection;
KW	human immunodeficiency; cancer; Th1 hyperfunction; anti-allergic;
KW	anti-inflammatory; antidiabetic; anti-lipemic; anti-infective; anti-HIV;
KW	cycostatic.; gene therapy; antisense therapy; ribozyme therapy; gene; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200296943-A1.
XX	
PD	05-DEC-2002.
XX	
PF	22-MAY-2002; 2002WO-JP004949.
XX	
PR	25-MAY-2001; 2001JP-00157043.
PR	30-AUG-2001; 2001JP-00260681.
PR	10-OCT-2001; 2001JP-00313175.
XX	
PA	(ASAH ) ASAHI KASEI KOGYO KK.
XX	
PI	Honda G, Matsuda A, Muramatsu S, Ishizawa K;
XX	
DR	WPI; 2003-140442/13.
DR	P-PSDB; ADG10694.
XX	
PT	STAT6-activating proteins and encoded genes, applicable in diagnosis of
PT	and developing drugs to treat allergic diseases, inflammations,
PT	autoimmune diseases, diabetes, hyperlipidemia, infections e.g. HIV, and
PT	cancer.
XX	
PS	Claim 4; SEQ ID NO 283; 2080pp; Japanese.

The invention relates to 242 human STAT6-activating proteins and cDNAs encoding them (ADG10411-ADG10894) and to sequences with 95% or more homology to the STAT6-activating proteins and their encoding nucleic acids. The invention also relates to recombinant vectors and host cells comprising a STAT6-activating protein-encoding nucleic acid; the recombinant production of a STAT6-activating protein; an antibody specific for a STAT6-activating protein; antisense oligonucleotides and ribozymes targeted to nucleic acids encoding a STAT6-activating protein; methods of screening for activators or inhibitors of STAT6-activating proteins; drug compositions comprising a modulator of STAT6-activating protein activity or expression; and methods of treating patients by administration of the drug compositions. The STAT6-activating proteins, nucleic acids encoding them, and modulators of their activity or expression are useful in the diagnosis and treatment of allergic diseases, inflammation, autoimmune diseases, diabetes, hyperlipidaemia, cancer, infections (e.g., HIV), cancer and disorders associated with Th1 hyperfunction. The present sequence is related to the invention.

Sequence 847 BP; 160 A; 312 C; 235 G; 140 T; 0 U; 0 Other

```

Alignment Scores:
Pred. NO.:      2.04e-24      Length:      847
Score:          261.50       Matches:     55
Percent Similarity: 65.18%   Conservative: 18
Best Local Similarity: 49.11% Mismatches:    30
Query Match:      40.92%     Indels:      10
DB:              10         Gaps:       1

US-10-657-740-1_COPY_51_173 (1-123) x ADG10693 (1-847)

Qy  7  LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp  26
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db  348 CTCAGCAGCGGGTCTCGGAGATCGGCACACTCGGACCGGTGGCGGTGTCCTGGAT  407

Qy  27  ValIysHisPheSerProGluAspLeuThrValIysValGlnAspAspPheValGluIle  46
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db  408 GTCAACCACTTCGCCCCCGAGCGAGCTGACGGTCAAGCAAGAGATGGCGTGTGGAGATC  467

Qy  47  HisGlyLysHisAsnGluAurGlnAspAspHisGlyIleSerArgGluPheHisArg  66
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db  468 ACCGCAAGCACGAGGAGCGCGCAGGACGAGCATGCTACATCTCCCGGTGCTTCACGCG  527

Qy  67  ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp  86
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db  528 AAATATACAGCTGCCCGCCCGGTGTGGACCCCAAGTTTCCTCCTGCTGCCCTGAG  587

Qy  87  GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGlu  106
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db  588 GGCACATGACCTGGA--GGCCCC-----CATGCCCAA  619

Qy  107 ArgAlaIleProValSerArgGluGluLysProThr  118
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db  620 GCTAGCCACGAGTCCCAACGAGATCAACCATCCAGT  655

```

RESULT 39	
ADQ86169	
ID	ADQ86169 standard; cDNA; 847 BP.
XX	
XX	
AC	ADQ86169;
XX	
XX	
DT	07-OCT-2004 (first entry)
XX	
XX	Human tumour-associated antigenic target (TAT) cDNA sequence #3041.
DE	
XX	
KW	human; tumour-associated antigenic target; TAT; cytostatic; gene therapy;
KW	cancer; cell proliferative disorder; gene; ss.
XX	
XX	
OS	Homo sapiens.
XX	
FN	WC2004060270-A2.
XX	
XX	
PD	22-JUL-2004.

XX 15-OCT-2003; 2003WO-US029126.  
 PF 18-OCT-2002; 2002US-0418988P.  
 PR (GETH ) GENENTECH INC.  
 PA (WUTD)/ WU T D.  
 PA (ZHOU)/ ZHOU Y.  
 XX Wu TD, Zhou Y;  
 XX WPI; 2004-534300/51.  
 DR New nucleic acid molecule and encoded polypeptide, for diagnosing,  
 XX preventing or treating cell proliferative disorders such as cancer.  
 PT Claim 1; SEQ ID NO 3041; 5504pp; English.  
 PS  
 CC The present invention describes an isolated tumour-associated antigenic  
 CC target (TAT) nucleic acid comprising: (a) any of 4622 nucleotide  
 CC sequences (see SEQ ID NO:1 to 4622); (b) the full-length coding region of  
 CC (a); (c) the complement of (a) or (b); (d) a sequence that has 80%  
 CC sequence identity to (a)-(c); or (e) a sequence that hybridises to (a)-  
 CC (c). Also described: (1) an expression vector comprising the above  
 CC nucleic acid; (2) a host cell comprising the above expression vector; (3)  
 CC a process for producing a polypeptide; (4) an isolated polypeptide  
 CC comprising: (a) an amino acid sequence encoded by any of the above  
 CC nucleotide sequences; (b) an amino acid sequence encoded by the full-  
 CC length coding region of the above nucleotide sequences; or (c) a sequence  
 CC having at least 80% identical to (a) or (b); (5) a chimeric polypeptide  
 CC comprising the above polypeptide fused to a heterologous polypeptide; (6)  
 CC an isolated antibody that binds to the above polypeptide; (7) a process  
 CC for producing the antibody; (8) an isolated oligopeptide that binds to  
 CC the above polypeptide; (9) a tumour-associated antigenic target (TAT)  
 CC binding organic molecule that binds to the above polypeptide; (10) a  
 CC composition of matter comprising the above (chimeric) polypeptide,  
 CC antibody, oligopeptide or TAT binding organic molecule, in combination  
 CC with a carrier; (11) an article of manufacture comprising a container and  
 CC the composition of matter contained within the container; (12) methods of  
 CC inhibiting the growth of a cell that expresses the above protein, where  
 CC the growth of the cell is at least in part dependent upon a growth  
 CC potentiating effect of the above protein; (13) a method of  
 CC therapeutically treating a mammal having a cancerous tumour comprising  
 CC cells that express the above protein; (14) a method of determining the  
 CC presence of a protein in a sample suspected of containing the protein  
 CC described above; (15) methods of diagnosing the presence of a tumour in a  
 CC mammal; (16) a method for treating or preventing a cell proliferative  
 CC disorder associated with increased expression or activity of the above  
 CC protein; and (17) a method of binding an antibody, oligopeptide or  
 CC organic molecule to a cell that expresses the protein described above.  
 CC The TAT sequences have cytostatic activities, and can be used in gene  
 CC therapy. The composition and methods are useful for diagnosing,  
 CC preventing or treating cancer. The composition is also used for preparing  
 CC a medicament for the therapeutic treatment or diagnostic detection of a  
 CC cell proliferative disorder or cancer. The present sequence represents a  
 CC human TAT cDNA sequence from the present invention.  
 XX  
 SQ Sequence 847 BP; 160 A; 313 C; 235 G; 139 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 2,04e-24 Length: 847  
 Score: 261.50 Matches: 55  
 Percent Similarity: 65.18% Conservative: 18  
 Best Local Similarity: 49.11% Mismatches: 30  
 Query Match: 40.92% Indels: 10  
 DB: 13 Gaps: 1  
 US-10-657-740-1\_COPY\_51\_173 (1-123) x ADQ86169 (1-847)  
 OY 7 LeuaspSerGlyIleSerGluValArgSerAspArgAspIlePheLeuAsp 26  
 DB 348 CTCAGACGGGGTCTGGAGATCGGCACACTCGGACCGCTGGCGTCTCCCTGGAT 407

OY 27 VallysHisPheSerProGluAspLeuThrVallysValGlnAspPheValGluIle 46  
 DB 408 GTCAACCACTTGGCCCGGAGCGTCAAGACCAAGGATGGCTGGTGGAGATC 467  
 OY 47 HisGlyIleHisAsnGluArgGlnAspHisGlyIleSerArgGluPheHisArg 66  
 DB 468 ACCGCAAGCAGCAGGAGCGGACGAGCATGCTACATCTCCCGGTGTTACGCGG 527  
 OY 67 ArgTyArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86  
 DB 528 AATATACAGCTGCCCGCGGTGGAGCCCAAGTTTCTCTCTCTCTCTCTCTCTCT 587  
 OY 87 GlyMetLeuThrPheCysGlyProIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106  
 DB 588 GGCACACTGACCGTGA-GGCCCC-----CATGCCCAA 619  
 OY 107 ArgAlaIleProValSerArgGluGluLysProThr 118  
 DB 620 GCTAGCCACGCGAGTCCCAACGAGATCACCATCCAGT 655  
 RESULT 40  
 ABQ60780  
 ID ABQ60780 standard; cDNA; 865 BP.  
 AC ABQ60780;  
 XX 02-AUG-2002 (first entry)  
 DE Human HSBP1 cDNA sequence SEQ ID NO:4480.  
 KW Human; colon cancer; cancer; tissue profiling; forensic; mapping;  
 KW genetic analysis; diagnostic; antisense therapy; gene; ss.  
 OS Homo sapiens.  
 XX WO2002229086-A2.  
 PD 11-APR-2002.  
 PF 02-OCT-2001; 2001WO-US030732.  
 PR 02-OCT-2000; 2000US-0237271P.  
 XX (FARB ) BAYER CORP.  
 PA Burges C, Astle JH, Carroll B, Catino TJ, Dwivedi P, Molino GA;  
 PI Thiagalingam A, Lewis ME;  
 PI WPI; 2002-426115/45.  
 DR P-PSDB; ABB78997.  
 XX New isolated nucleic acid that is differentially expressed in cancer  
 PT tissues useful for determining the presence of colon cancer in a cell or  
 PT tissue type, and in antisense therapy.  
 XX Claim 1; Fig 2; 796pp; English.  
 CC ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differentially  
 CC expressed in cancer tissues. ABB78993 to ABB79004 represent proteins  
 CC encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) can be  
 CC used in antisense therapy. An antibody immunoreactive with a polypeptide  
 CC encoded by (I) is useful for detecting cancer in a patient sample, and  
 CC for detecting the presence or absence of a polynucleotide encoded by a  
 CC nucleic acid which hybridises to (I) in a cell. A probe/primer derived  
 CC from (I) can be used for determining the presence of a nucleic acid which  
 CC hybridises to (I), and for determining the phenotype of cells in a sample  
 CC of cells from a patient. (I) is useful for determining the presence of  
 CC colon cancer in a cell or tissue type, for determining the presence or  
 CC state of other type of cancer, in antisense therapy, to generate  
 CC macroarrays on a solid surface, to identify a chromosome on which the  
 CC corresponding gene resides, and in tissue profiling, forensics, genetic  
 CC analysis, mapping and diagnostic applications. (I) can be used to raise  
 CC antibodies, and to screen for peptide analogues and antagonists



```
XX SQ Sequence 865 BP; 178 A; 312 C; 235 G; 140 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 2.1e-24 Length: 865
Score: 261.50 Matches: 55
Percent Similarity: 65.18% Conservative: 18
Best Local Similarity: 49.11% Mismatches: 30
Query Match: 40.92% Indels: 10
DB: 13 Gaps: 1
US-10-657-740-1_COPY_51_173 (1-123) x ABQ60780 (1-865)
QY 7 LeuaspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuasp 26
Db 348 CTCACAGCGGGGTCTCGAGATCCGGCACACTGCGGACCGCTGCGGTGCTCCCTGGAT 407
QY 27 VallysHisPheSerProGluAspLeuThrVallysValGlnAspPheValGluile 46
Db 408 GTCAACCACTTGGCCCGGACGAGCTGACGCTCAAGACCAAGGATGGCGTGGAGATC 467
QY 47 HisGlyLysHisAsnGluArgGlnAspHisGlyTyrlleSerArgGluPheHisArg 66
Db 468 ACCGCAAGCAGCAGGAGCGGACGAGCATGCTACATCTCCGGTGTTCACGCGG 527
QY 67 ArgTyArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86
Db 528 AAATACACGCTGCCCGCGGTGTGGACCCACCCCAAGTTTCTCCTCCTGTCCCTGAG 587
QY 87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAlaThrHisAlaGlu 106
Db 588 GGCACACTGACCGTGA-GGCCCC-----CATGCCCAA 619
QY 107 ArgAlaIleProValSerArgGluLysProThr 118
Db 620 GCTAGCCAGCAGTCCACGAGATCACCATCCCACT 655
RESULT 41
ADR24765
ID ADR24765 standard; DNA; 865 BP.
XX AC ADR24765;
XX DT 21-OCT-2004 (first entry)
XX DE Breast cancer prognosis marker #626.
XX KW ds; breast cancer; prognosis; gene expression; diagnosis.
XX OS Homo sapiens.
XX PN WO2004065545-A2.
XX PD 05-AUG-2004.
XX PF 15-JAN-2004; 2004WO-US001100.
XX PR 15-JAN-2003; 2003US-00342887.
XX PA (ROSE-) ROSETTA INPHARMATICS LLC.
XX PA (NECA-) NETHERLANDS CANCER INST.
XX FI Van't Veer LJ, He Y;
XX KW MPI; 2004-593473/57.
XX DR
XX PT Classifying a breast cancer patient according to prognosis comprises
XX PT determining the similarity between the level of expression of each of
XX PT five genes in a cell sample taken from patient, to control levels.
XX PS Disclosure; SEQ ID NO 626; 226pp; English.
XX CC The invention relates to a method of classifying a breast cancer patient
```

```
CC according to prognosis by determining the similarity between the level of
CC expression of each of five genes for which markers are listed in the
CC specification, in a cell sample taken from the breast cancer patient, to
CC control levels of expression for each respective five genes to obtain a
CC patient similarity value. The methods are useful for classifying a breast
CC cancer patient according to prognosis, kits and computer program products
CC are useful for data analysis using the diagnostic, prognostic and
CC statistical methods of the invention. This sequence corresponds to a
CC marker used in the method of the invention.
XX SQ Sequence 865 BP; 178 A; 312 C; 235 G; 140 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 2.1e-24 Length: 865
Score: 261.50 Matches: 55
Percent Similarity: 65.18% Conservative: 18
Best Local Similarity: 49.11% Mismatches: 30
Query Match: 40.92% Indels: 10
DB: 13 Gaps: 1
US-10-657-740-1_COPY_51_173 (1-123) x ADR24765 (1-865)
QY 7 LeuaspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuasp 26
Db 348 CTCACAGCGGGGTCTCGAGATCCGGCACACTGCGGACCGCTGCGGTGCTCCCTGGAT 407
QY 27 VallysHisPheSerProGluAspLeuThrVallysValGlnAspPheValGluile 46
Db 408 GTCAACCACTTGGCCCGGACGAGCTGACGCTCAAGACCAAGGATGGCGTGGAGATC 467
QY 47 HisGlyLysHisAsnGluArgGlnAspHisGlyTyrlleSerArgGluPheHisArg 66
Db 468 ACCGCAAGCAGCAGGAGCGGACGAGCATGCTACATCTCCGGTGTTCACGCGG 527
QY 67 ArgTyArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86
Db 528 AAATACACGCTGCCCGCGGTGTGGACCCACCCCAAGTTTCTCCTCCTGTCCCTGAG 587
QY 87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAlaThrHisAlaGlu 106
Db 588 GGCACACTGACCGTGA-GGCCCC-----CATGCCCAA 619
QY 107 ArgAlaIleProValSerArgGluLysProThr 118
Db 620 GCTAGCCAGCAGTCCACGAGATCACCATCCCACT 655
RESULT 42
ACN38656
ID ACN38656 standard; cDNA; 865 BP.
XX AC ACN38656;
XX DT 18-NOV-2004 (first entry)
XX DE Tumour-associated antigenic target (TAT) cDNA DNA304710, SEQ ID NO.2308.
XX KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;
XX KW tumour; diagnosis; cell proliferative disorder; breast cancer;
XX KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;
XX KW central nervous system cancer; bladder cancer; pancreatic cancer;
XX KW cervical cancer; melanoma; leukaemia; hybridisation probe;
XX KW chromosome identification; chromosome mapping; gene mapping;
XX KW gene therapy; cytostatic; gene; ss.
XX OS Homo sapiens.
XX PN WO2004030615-A2.
XX PD 15-APR-2004.
XX PF 29-SEP-2003; 2003WO-US028547.
XX PR 02-OCT-2002; 2002US-0414971P.
```

XX (GETH ) GENENTECH INC.  
 XX Wu TD, Zhang Z, Zhou Y;  
 XX WPI; 2004-347921/32.  
 XX P-PSDB; ABM0896.  
 XX New tumor-associated antigenic target polypeptides and nucleic acids,  
 PT useful in preparing a medicament for treating or detecting a  
 PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or  
 PT prostate cancer or tumor.  
 XX  
 PS Claim 1; SEQ ID NO 2308; 7273pp; English.  
 PS  
 CC The invention relates to human tumour-associated antigenic target (TAT)  
 CC polypeptides, and their related nucleic acids. The TAT polypeptides are  
 CC overexpressed in cancer tissues compared to normal tissues, and may thus  
 CC serve as effective targets for the diagnosis and treatment of cancer in  
 CC mammals. The invention also relates to nucleic acid and polypeptide  
 CC sequences at least 80% identical to the TAT nucleic acids and  
 CC polypeptides; expression vectors and host cells comprising a TAT nucleic  
 CC acid; an antibody specific for a TAT polypeptide; a peptide or organic  
 CC molecule which binds to a TAT polypeptide; fusion proteins comprising a  
 CC TAT polypeptide; and methods and compositions for the treatment or  
 CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,  
 CC antibodies, antagonists, binding molecules and compositions are useful  
 CC for diagnosing or treating a cell proliferative disorder associated with  
 CC increased TAT expression, particularly cancers such as breast cancer,  
 CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder  
 CC cancer, pancreatic cancer, cervical cancer, cancers of the central  
 CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be  
 CC used as hybridisation probes, in chromosome and gene mapping, in  
 CC chromosome identification and in gene therapy. The present sequence  
 CC represents a TAT nucleic acid of the invention  
 XX  
 XX Sequence 865 BP; 178 A; 312 C; 235 G; 140 T; 0 U; 0 Other;  
 XX

Alignment Scores:  
 Pred. No.: 2,1e-24 Length: 865  
 Score: 261.50 Matches: 55  
 Percent Similarity: 65.18% Conservative: 18  
 Best Local Similarity: 49.11% Mismatches: 30  
 Query Match: 40.92% Indels: 10  
 DB: 13 Gaps: 1

US-10-657-740-1\_COPY\_51\_173 (1-123) x ACN38656 (1-865)

QY 7 LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26  
 Db 348 CTCAGCAGCGGGTCTCGGAGATCGGCACACTGGCGACCGGTGGCGTGTCTCGAT 407  
 QY 27 ValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGluLeu 46  
 Db 408 GTCAACCACTTCGCCCGGACGAGCTGACGGTCAAGACACAGAGATGGCGTGTGGAGATC 467  
 QY 47 HisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArg 66  
 Db 468 ACCGCGACGACGAGGAGCGGCGAGCAGCATGCTACTCTCCGTGCTTACGCGG 527  
 QY 67 ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86  
 Db 528 AATACACAGCTGCCCGCGGTGTGGACCCCAAGTTCTCTCCCTCTGTCCTCCCTGAG 587  
 QY 87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106  
 Db 588 GGCACACTGACCGTGA-GGCCCC-----CATGCCCAA 619  
 QY 107 ArgAlaIleProValSerArgGluLysProThr 118  
 Db 620 GCTAGCCACGAGTCCACGAGATCACCATCCCACT 655

RESULT 43

ADP23292  
 ID ADP23292 standard; cDNA; 865 BP.  
 XX  
 AC ADP23292;  
 XX  
 DT 18-NOV-2004 (first entry)  
 XX  
 XX PRO polypeptide encoding cDNA SEQ ID NO:386.  
 DE  
 XX ss; gene; PRO; antiinflammatory; antiarthritic; antirheumatic;  
 KW immunosuppressive; osteopathic; antidiabetic; dermatological;  
 KW antipsoriatic; antiallergic; antiasthmatic; hepatotropic; respiratory;  
 KW gene therapy; immune system.  
 XX  
 OS Unidentified.  
 XX  
 XX WO2004041170-A2.  
 FN  
 XX 21-MAY-2004.  
 PD  
 XX 30-OCT-2003; 2003WO-US034312.  
 PF  
 XX 01-NOV-2002; 2002US-0423394P.  
 PR  
 XX (GETH ) GENENTECH INC.  
 PA  
 XX Clark H, Schoenfeld J, Van Lookeren M, Williams PM, Wood WI;  
 PI Wu TD;  
 XX  
 XX WPI; 2004-419628/39.  
 DR P-PSDB; ADP23293.  
 XX  
 XX New PRO polypeptides and polynucleotides, useful for treating e.g.  
 PT erythematous, rheumatoid arthritis, diabetes mellitus, immune-mediated  
 PT renal disease, or demyelinating diseases of the central or peripheral  
 PT nervous system.  
 XX  
 PS Claim 1; SEQ ID NO 386; 2940pp; English.  
 XX  
 CC The invention relates to a novel isolated nucleic acid and the PRO  
 CC polypeptide encoded by it. A protein of the invention has  
 CC antiinflammatory, antiarthritic, antirheumatic, immunosuppressive,  
 CC osteopathic, antidiabetic, dermatological, antipsoriatic, antiallergic,  
 CC antiasthmatic, hepatotropic, and respiratory activity. A polynucleotide  
 CC of the invention may have a use in gene therapy. The PRO polypeptide, its  
 CC agonist, antagonist, or antibody that specifically binds to the  
 CC polypeptide is useful for treating an immune related disorder such as  
 CC systemic lupus erythematosus, rheumatoid arthritis, osteoarthritis,  
 CC juvenile chronic arthritis, a spondyloarthropathy, systemic sclerosis, an  
 CC idiopathic inflammatory myopathy, Sjogren's syndrome, systemic  
 CC vasculitis, sarcoidosis, autoimmune haemolytic anaemia, autoimmune  
 CC thrombocytopenia, thyroiditis, diabetes mellitus, immune-mediated renal  
 CC disease, a demyelinating disease of the central or peripheral nervous  
 CC system, idiopathic demyelinating polyneuropathy, Guillain-Barre syndrome,  
 CC a chronic inflammatory demyelinating polyneuropathy, a hepatobiliary  
 CC disease, infectious or autoimmune chronic active hepatitis, primary  
 CC biliary cirrhosis, granulomatous hepatitis, sclerosing cholangitis,  
 CC inflammatory bowel disease, gluten-sensitive enteropathy, Whipple's  
 CC disease, an autoimmune or immune-mediated skin disease, a bullous skin  
 CC disease, erythema multiforme, contact dermatitis, psoriasis, an allergic  
 CC disease, asthma, allergic rhinitis, atopic dermatitis, food  
 CC hypersensitivity, urticaria, an immunologic disease of the lung,  
 CC eosinophilic pneumonia, idiopathic pulmonary fibrosis, hypersensitivity  
 CC pneumonitis, a transplantation associated disease, graft rejection or  
 CC graft-versus-host disease. The present sequence encodes a PRO protein of  
 CC the invention.  
 XX  
 XX Sequence 865 BP; 178 A; 312 C; 235 G; 140 T; 0 U; 0 Other;  
 XX

Alignment Scores:  
 Pred. No.: 2,1e-24 Length: 865  
 Score: 261.50 Matches: 55  
 Percent Similarity: 65.18% Conservative: 18

Best Local Similarity: 49.11% Mismatches: 30  
 Query Match: 40.92% Indels: 10  
 DB: 13 Gaps: 1

US-10-657-740-1\_COPY\_51\_173 (1-123) x ADP23292 (1-865)

Qy 7 LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26  
 Db 348 CTCAGCAGCGGGTCTCGAGATCCGGACACTGCGGACCGCTGGCGGTGTCTCGAT 407

Qy 27 VallysHisPheSerProGluAspLeuThrVallysValGlnAspAspPheValGluIle 46  
 Db 408 GTCAACCACTTCGCCCGGACGAGTGAAGTCAAGACCAAGAGTGGCGTGTGGAGATC 467

Qy 47 HisGlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArg 66  
 Db 468 ACCGCAAGCAGCAGGAGCGGACGAGCAGTGTCTCATCTCCCGGTGTTTCACGCGG 527

Qy 67 ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86  
 Db 528 AATACACGCTGCCCGGCGGTGGAGCCCAAGTTCTCTCTCTGTCCCTGTCCCTGAG 587

Qy 87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106  
 Db 588 GGCACACTGACCGTGA-GGCCCC-----CATGCCCAA 619

Qy 107 ArgAlaIleProValSerArgGluGluLysProThr 118  
 Db 620 GCTAGCCACGAGTCCCAACGAGATCACCATCCCACT 655

RESULT 44  
 ABN97370  
 ID ABN97370 standard; DNA; 1231 BP.  
 XX AC ABN97370;  
 XX DT 13-AUG-2002 (first entry)  
 XX DE Gene #3868 used to diagnose liver cancer.  
 KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;  
 KW metastatic liver tumour; cytostatic; expression profile; disease state;  
 KW disease progression; drug toxicity; drug efficacy; drug metabolism.  
 XX OS Homo sapiens.  
 XX WO200229103-A2.  
 XX PD 11-APR-2002.  
 XX PF 02-OCT-2001; 2001WO-US030589.  
 XX PR 02-OCT-2000; 2000US-0237054P.  
 XX PA (GENE-) GENE LOGIC INC.  
 XX PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;  
 XX WPI; 2002-426119/45.  
 XX PT Diagnosing and detecting the progression of liver cancer, hepatocellular  
 PT carcinoma or metastatic liver tumor in a patient, involves detecting the  
 PT level of expression of two or more genes in a liver tissue sample.  
 XX PS Claim 1; SEQ ID NO 3868; 298pp; English.  
 XX CC The invention relates to a novel method for diagnosing and detecting the  
 CC progression of liver cancer, hepatocellular carcinoma or metastatic liver  
 CC tumour in a patient, and differentiating metastatic liver cancer from  
 CC hepatocellular carcinoma in a patient, involving detecting the level of  
 CC expression of two or more genes represented in ABN93503-ABN97455 in a  
 CC tissue sample. The method of the invention has hepatotropic, and  
 CC cytostatic activity. The method is useful for diagnosing and detecting

CC the progression of liver cancer, hepatocellular carcinoma and metastatic  
 CC liver carcinoma in a patient. The method is useful for identifying  
 CC expression profiles which serve as useful diagnostic markers as well as  
 CC markers that can be used to monitor disease states, disease progression,  
 CC drug toxicity, drug efficacy and drug metabolism. Note: The sequence data  
 CC for this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 1231 BP; 245 A; 396 C; 321 G; 269 T; 0 U; 0 Other;

Alignment Scores: 3.49e-24 Length: 1231  
 Pred. No.: 261.50 Matches: 55  
 Score: 65.18% Conservative: 18  
 Percent Similarity: 49.11% Mismatches: 30  
 Best Local Similarity: 40.92% Indels: 10  
 Query Match: 6 Gaps: 1  
 DB: 1

US-10-657-740-1\_COPY\_51\_173 (1-123) x ABN97370 (1-1231)

Qy 7 LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26  
 Db 732 CTCAGCAGCGGGTCTCGAGATCCGGACACTGCGGACCGCTGGCGGTGTCTCGAT 791

Qy 27 VallysHisPheSerProGluAspLeuThrVallysValGlnAspAspPheValGluIle 46  
 Db 792 GTCAACCACTTCGCCCGGACGAGTGAAGTCAAGACCAAGAGTGGCGTGTGGAGATC 851

Qy 47 HisGlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArg 66  
 Db 852 ACCGCAAGCAGCAGGAGCGGACGAGCATGCTATCTCCCGGTGTTTCACGCGG 911

Qy 67 ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86  
 Db 912 AATACACGCTGCCCGGCGGTGGAGCCCAAGTTCTCTCTCTGTCCCTGTCCCTGAG 971

Qy 87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106  
 Db 972 GGCACACTGACCGTGA-GGCCCC-----CATGCCCAA 1003

Qy 107 ArgAlaIleProValSerArgGluGluLysProThr 118  
 Db 1004 GCTAGCCACGAGTCCCAACGAGATCACCATCCCACT 1039

RESULT 45  
 ADD70998  
 ID ADD70998 standard; DNA; 1231 BP.  
 XX AC ADD70998;  
 XX DT 15-JAN-2004 (first entry)  
 XX DE Human heat shock 27kD protein 1 gene SEQ ID NO:2.  
 XX KW liver cancer; chronic hepatitis; cirrhosis; liver disease; hepatotropic;  
 KW cytostatic; gene therapy; human; gene; ds.  
 XX OS Homo sapiens.  
 XX WO2003061564-A2.  
 XX PF 31-JUL-2003.  
 XX PR 20-DEC-2002; 2002WO-US040718.  
 XX PR 21-DEC-2001; 2001US-0341815P.  
 XX PR 31-DEC-2001; 2001US-0343185P.  
 XX PA (GENE-) GENE LOGIC INC.  
 XX PA (JGHI-) LG BIOMEDICAL INST.  
 XX PI Koh SS, Liu Q, Chung H, Zeng W, Lee B, Yeramilli S, Song SY;

XX DR WPI; 2003-663343/62.  
 XX PT Diagnosing liver cancer cells, useful for treating liver cancer  
 XX PT associated with chronic hepatitis or cirrhosis comprises detecting the  
 XX PT level of expression in a tissue sample of one or more genes associated  
 XX PT with cancerous liver tissues.  
 XX PS Claim 1; SEQ ID NO 2; 176pp; English.  
 XX CC The present invention describes a method for diagnosing liver cancer  
 CC cells comprising detecting the level of expression in a tissue sample of  
 CC one or more genes given in the specification (see ADD70997 to ADD71105),  
 CC where differential expression of the genes is indicative of liver cancer.  
 CC Also described: (1) detecting the progression of liver cancer in a  
 CC patient; (2) monitoring the treatment of a patient with liver cancer; (3)  
 CC treating a patient with liver cancer; (4) typing a liver disease in a  
 CC patient; (5) detecting the presence or progression of liver cancer in a  
 CC patient with chronic hepatitis or cirrhosis; (6) differentiating liver  
 CC cancer related to chronic hepatitis from liver cancer related to  
 CC cirrhosis; (7) screening for an agent capable of modulating the onset or  
 CC progression of liver cancer; (8) a composition comprising at least two  
 CC oligonucleotides comprising a sequence that specifically hybridises to  
 CC any of the genes; (9) a solid support comprising the at least two  
 CC oligonucleotides; (10) a computer system comprising a database containing  
 CC information identifying the level in liver tissue of a set of genes; (11)  
 CC a method for using the computer system to present information identifying  
 CC the expression level in tissue or cell of any of the genes; and (12) a  
 CC therapeutic agent for slowing or halting the progression of liver cancer.  
 CC The methods are useful for treating liver cancer associated with chronic  
 CC hepatitis or cirrhosis. The present sequence represents a specifically  
 CC claimed human gene sequence which is used in the exemplification of the  
 CC present invention.

XX SQ Sequence 1231 BP; 245 A; 396 C; 321 G; 269 T; 0 U; 0 Other;

# Alignment Scores:

Pred. No.: 3,49e-24 Length: 1231  
 Score: 261.50 Matches: 55  
 Percent Similarity: 65.18% Conservative: 18  
 Best Local Similarity: 49.11% Mismatches: 30  
 Query Match: 40.92% Indels: 10  
 DB: 10 Gaps: 1

US-10-657-740-1\_COPY\_51\_173 (1-123) x ADD70998 (1-1231)

QY 7 LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26  
 DB 732 CTCAGCAGCGGGTCTCGGAGATCGGCACACTGCGACCGTGGCGGTGTCTCTGGAT 791  
 QY 27 ValLysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGluLe 46  
 DB 792 GTCAACCACTTCGCCCGGACGAGTGGTCAAGACCAAGATGGGTGGTGAGATC 851  
 QY 47 HisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArg 66  
 DB 852 ACCGCGACGACGAGCGGCGGACGACGACGACGACGACGACGACGACGACGACG 911  
 QY 67 ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86  
 DB 912 AAATACAGCTGCCCGCGGTGTGACGCCACCCACCAAGTTCTCTCTCTCTCTCTCTGAG 971  
 QY 87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106  
 DB 972 GCACACTGACCGTGA-GGCCCC-----CATGCCCAA 1003  
 QY 107 ArgAlaIleProValSerArgGluGluLysProThr 118  
 DB 1004 GCTAGCCAGCGATCCACGAGATCACCATCCAGT 1039

RESULT 46

AAA93441

ID AAA93441 standard; DNA; 1380 BP.

XX AC AAA93441;  
 XX DT 15-SEP-2003 (revised)  
 XX DT 10-JAN-2001 (first entry)  
 XX DE GFP-HSP27 fusion gene, SEQ ID NO:169.  
 XX KW Bioreactor protein; fusion protein; recognition site;  
 KW cellular targeting sequence; cellular localisation; fluorescent protein;  
 KW protease activity detection; toxin detection; cellular stress detection;  
 KW drug discovery; cell based screening; ds.  
 OS Aequorea victoria.  
 OS Mammalia:  
 OS Chimeric.  
 XX PN WO200050872-A2.  
 XX 31-AUG-2000.  
 XX 25-FEB-2000; 2000WO-US004794.  
 XX 26-FEB-1999; 99US-0122152P.  
 XX 08-MAR-1999; 99US-0123399P.  
 XX 12-JUL-1999; 99US-00352171.  
 XX (CELL-) CELLOMICS INC.  
 XX Giuliano KA, Kapur R;  
 XX WPI; 2000-594086/56.  
 XX P-PSDB; AAB22936.  
 XX Automated cell-based characterization of toxin by contacting cells  
 XX containing luminescent reporter molecules with test substance and  
 XX analyzing optically.  
 XX Example 11; Page 300-302; 336pp; English.  
 XX The invention relates to systems, methods and reagents for cell-based  
 XX screening or detection of compounds which affect particular biological  
 XX functions. The methods of the invention utilise fluorescent bioreactor  
 XX molecules which, when acted on by a compound of interest, cause an  
 XX alteration in the cellular distribution of at least the fluorescent  
 XX moiety. In one embodiment, the biosensors comprise heat shock proteins  
 XX (HSPs) fused to a fluorescent protein (e.g., jellyfish green fluorescent  
 XX protein (GFP), or derivatives thereof). Such biosensors are located in  
 XX the cytoplasm, but on stress activation translocate to the nucleus. In  
 XX another embodiment bioreactor proteins can be used to detect protease  
 XX activity. Such protease bioreactor fusion proteins comprise one or more  
 XX fluorescent proteins; a recognition signal which is cleaved by the  
 XX protease; and at least one cellular localisation signal. The latter two  
 XX components may be components of a single protein which is acted upon by  
 XX the protease, or may be from heterologous sources. Due to the  
 XX localisation signal, the bioreactor protein is localised to a particular  
 XX region of the cell. Once acted on by the protease of interest, the  
 XX fluorescent protein is cleaved from the localisation sequence, and is  
 XX free to migrate to other locations within the cell. The presence of a  
 XX second localisation signal attached to the fluorescent protein enables  
 XX the fluorescent protein to be directed to a different cellular  
 XX compartment after cleavage of the protease recognition sequence. The  
 XX change in distribution of the fluorescent protein can be detected using  
 XX imaging methods with a high degree of spatial resolution. The methods and  
 XX biosensors of the invention can be used to investigate a wide range of  
 XX cellular activities and to screen compounds which modulate these  
 XX activities. Biosensors containing a recognition site for caspase, for  
 XX example, may be used for the screening of compounds which modulate  
 XX apoptosis, while biosensors containing other protease recognition sites  
 XX may be used for the detection of proteolytic toxins (such as anthrax  
 XX lethal factor). The method provides improved target validation and  
 XX candidate compound optimisation by combining many cell screening formats  
 XX with fluorescence-based molecular reagents and computer-based feature



AC ABK63074;  
 DT 18-JUN-2002 (first entry)  
 DE Rat sequence differentially expressed in response to a hepatotoxin #981.  
 XX Rat; sg; hepatotoxin; expressed sequence tag; EST; drug screening;  
 KW differential expression; centrilobular necrosis; steatosis.  
 XX Rattus norvegicus.  
 OS  
 XX  
 XX  
 PN WO200210453-A2.  
 XX  
 PD 07-FEB-2002.  
 XX  
 XX  
 PF 30-JUL-2001; 2001WO-US023872.  
 XX  
 PR 31-JUL-2000; 2000US-0222040P.  
 PR 02-NOV-2000; 2000US-0244880P.  
 PR 11-MAY-2001; 2001US-0290029P.  
 PR 15-MAY-2001; 2001US-0290645P.  
 PR 22-MAY-2001; 2001US-0292336P.  
 PR 06-JUN-2001; 2001US-02925798P.  
 PR 13-JUN-2001; 2001US-0297457P.  
 PR 19-JUN-2001; 2001US-0298884P.  
 PR 09-JUL-2001; 2001US-0303459P.  
 XX  
 PA (GENE-) GENE LOGIC INC.  
 XX  
 XX Mendrick D, Porter MW, Johnson KR, Castle AL, Elashoff MR;  
 XX WPI; 2002-241625/29.  
 DR  
 XX  
 XX Predicting toxic effects of compounds or the progression of these toxic  
 PT effects by determining the changes in gene expression in tissues or cells  
 PT exposed to the toxin and comparing these to gene expression in unexposed  
 PT tissues or cells.  
 XX  
 PS Claim 1; SEQ ID NO 981; 239pp; English.  
 XX  
 CC The invention relates to methods for predicting toxic effects of  
 CC compounds or the progression of these toxic effects by determining the  
 CC global changes in gene expression in tissues or cells exposed to the  
 CC toxin and comparing these to gene expression in unexposed tissues or  
 CC cells. Also included are methods of predicting at least one toxic effect  
 CC of a compound or progression of a toxic effect, preferably the  
 CC hepatotoxicity of a compound, comprising detecting the level of  
 CC expression in a tissue or cell sample exposed to the compound of two or  
 CC more genes listed in the specification, where differential expression of  
 CC the genes is indicative of at least one toxic effect or progression. The  
 CC method can also be used to identify an agent which modulates the toxic  
 CC response and predict cellular pathways that a compound modulates in a  
 CC cell. The methods utilise a set of at least two probes (on a solid  
 CC support in kit form), where each of the probes comprises a sequence that  
 CC specifically hybridises to a gene listed in the specification, a computer  
 CC system comprising a database containing information identifying the  
 CC expression level in a tissue or cell sample exposed to a hepatotoxin of a  
 CC set of genes comprising at least two genes listed in the specification,  
 CC and a user interface to view the information used to present information  
 CC identifying the expression level in a tissue or cell of at least one gene  
 CC listed in the specification. The method is useful for elucidating global  
 CC changes in gene expression and for identifying toxicity markers in  
 CC tissues or cell exposed to a known toxin. The genes may be used as  
 CC toxicity markers in drug screening and toxicity assays. The genes and  
 CC gene expression information may be used as diagnostic markers for the  
 CC prediction or identification of the physiological state of tissue or cell  
 CC sample that has been exposed to a compound or agent. Hepatotoxicity is  
 CC characterised by centrilobular necrosis and steatosis. The present  
 CC sequence is an expressed sequence tag (EST) or cDNA derived from a gene  
 CC which is differentially expressed in response to a hepatotoxic agent  
 XX  
 SQ Sequence 604 BP; 124 A; 156 C; 199 G; 125 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 1,46e-24 Length: 604  
 Score: 261.00 Matches: 57  
 Percent Similarity: 65.18% Conservative: 16  
 Best Local Similarity: 50.89% Mismatches: 33  
 Query Match: 40.85% Indels: 6  
 DB: 6 Gaps: 2

US-10-657-740-1\_COPY\_51\_173 (1-123) x ABK63074 (1-604)

QY 7 LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26  
 DB 492 CTCAGTAGCGGTGCTCAGAGATCCGACAGACGGCGGATCGCTGGCGGTGCTCCCTGGAC 433  
 QY 27 ValIysHisPheSerProGluAspLeuThrValIysValGlnAspAspPheValGluIle 46  
 DB 432 GTCAACCACTTCGCTCTCAGAGAGCTCACATTAAAGACCAAGGAGCGGTGGTGGAGATC 373  
 QY 47 HisGlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArg 66  
 DB 372 ACTGGCAAGCAGCAAGAAAGGAGGATCAACATGGCTACATCTCTCGGTGCTTCACCCGG 313  
 QY 67 ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86  
 DB 312 AAATACAGCTCCCTCCAGGTGGACCCACCTTGGTGTCTCTCTCTCTCTCTCTCTCTCTG 253  
 QY 87 GlyMetLeuThrPheCysGly-----ProLysIleGlnThrGlyLeuAspAlaThrHis 104  
 DB 252 GGCACTCAGCGTGAAGGCTCCGCTCCCAAGACGACTCA-----CAATCA 205  
 QY 105 AlaGluArgAlaIleProValSerArgGluGluLys 116  
 DB 204 GCGGAGATCACTTTCGGTCACTTTCGAGGCGCGT 169

## RESULT 49

ADP72259/C  
 ID ADP72259 standard; DNA; 604 BP.

XX AC ADP72259;  
 XX DT 26-AUG-2004 (first entry)  
 XX DE Renal toxin progression gene marker #848.

XX KW ds; toxic effect; gene expression profile; kidney tissue;  
 XX KW differential gene expression; toxicity progression; toxicity marker;  
 XX KW drug screening; toxicity assay; kidney pathology; nephritis;  
 XX KW kidney necrosis; glomerular injury; tubular injury;  
 XX KW focal segmental glomerulosclerosis.  
 XX OS Rattus norvegicus.  
 XX PN WO2004048598-A2.  
 XX PD 10-JUN-2004.  
 XX PF 24-NOV-2003; 2003WO-US037556.  
 XX PR 22-NOV-2002; 2002US-00301856.  
 XX PA (GENE-) GENE LOGIC INC.

XX PI Mendrick DL, Porter MW, Johnson KR, Castle A, Higgs B;  
 XX PI Elashoff M;  
 XX DR WPI; 2004-460771/43.

XX PT Predicting (the progression of) a toxic effect of a compound, for  
 PT monitoring the progression of renal disease states, comprises preparing a  
 PT gene expression profile of a kidney tissue or cell sample exposed to the  
 PT compound.

PS Claim 11; SEQ ID NO 848; 266pp; English.

XX The invention relates to a method of predicting (the progression of) a  
CC toxic effect of a compound by preparing a gene expression profile of a  
CC kidney tissue or cell sample exposed to the compound and comparing the  
CC gene expression profile to a database, or detecting the level of gene(s)  
CC expression in a tissue or cell sample exposed to the compound, where  
CC differential gene expression compared to a control indicates a toxic  
CC effect (toxicity progression). The method is useful for predicting (the  
CC progression of) at least one toxic effect of a compound. The genes are  
CC useful as toxicity markers in drug screening and toxicity assays. The  
CC methods are useful for predicting the likelihood that a compound or test  
CC agent will induce various specific kidney pathologies, such as nephritis,  
CC kidney necrosis, glomerular and tubular injury, or focal segmental  
CC glomerulosclerosis. The methods are useful for determining the similarity  
CC of a toxic response to one or more individual compounds and for  
CC predicting or elucidating the potential cellular pathways influenced,  
CC induced or modulated by the compound or test agent. The kit is useful for  
CC predicting or modelling the toxic response of a test compound, for  
CC monitoring the progression of renal disease states, for identifying genes  
CC that show promise as new drug targets and for screening known and newly  
CC designed drugs. This sequence corresponds to a gene marker used in the  
CC method of the invention. (Note: The sequence data for this patent did not  
CC form part of the printed specification, but was obtained in electronic  
CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences).

XX Sequence 604 BP; 124 A; 156 C; 199 G; 125 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 1.46e-24 Length: 604  
Score: 261.00 Matches: 57  
Percent Similarity: 65.18% Conservative: 16  
Best Local Similarity: 50.89% Mismatches: 33  
Query Match: 40.85% Indels: 6  
DB: 12 Gaps: 2

US-10-657-740-1\_COPY\_51\_173 (1-123) x ADP72259 (1-604)

Qy 7 LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26  
Db 492 CTCAGTAGGGTGTCTCAGAGATCCGACGCGCGATCGTGGCGGTGTCTCCCTGGAC 433  
Qy 27 VallysHisPheSerProGluAspLeuThrVallysValGlnAspAspPheValGlu 46  
Db 432 GTCAACCACTTCGCTCCGTAGAGGTCACAGTTAAGACCAAGAGGCGGTGTGGAGATC 373  
Qy 47 HisGlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArg 66  
Db 372 ACTGCACAGCAGCAGAAAGGAGGATGACATGCTACATCTCTCGGTGCTTCAACCGG 313  
Qy 67 ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86  
Db 312 AAATACACGCTCCCTCCAGGTGTGAGACCCACCTTGTGTCTCTTCTCCCTGCTCCG 253  
Qy 87 GlyMetLeuThrPheCysGly-----ProLysIleGlnThrGlyLeuAspAlaThrHis 104  
Db 252 GGCACACTCAGCGGTGAAGGCTCCGCTGCCAAAGCAGTCACA-----CAATCA 205  
Qy 105 AlaGluArgAlaIleProValSerArgGluGluLys 116  
Db 204 GCGGAGATCACCATTTCGGTCTACTTTCGAGGCCGT 169

RESULT 50

ABV94670

ID ABV94670 standard; cDNA; 599 BP.

XX AC ABV94670;

XX DT 14-JAN-2003 (first entry)

XX DE Human pancreatic cancer expressed cDNA SEQ ID NO 43.

XX KW Human; pancreas; cancer; gene therapy; vaccine; immunostimulant;

XX cytotostatic; tumour; gene; ss.

XX OS. Homo sapiens.  
XX PN WO200260317-A2.  
XX PD 08-AUG-2002.

XX PF 30-JAN-2002; 2002WO-US002781.  
XX PR 30-JAN-2001; 2001US-0265305P.  
XX PR 31-JAN-2001; 2001US-0265682P.  
XX PR 09-FEB-2001; 2001US-0267568P.  
XX PR 21-MAR-2001; 2001US-0278651P.  
XX PR 28-APR-2001; 2001US-0287112P.  
XX PR 16-MAY-2001; 2001US-0291631P.  
XX PR 12-JUL-2001; 2001US-0305484P.  
XX PR 20-AUG-2001; 2001US-0313999P.  
XX PR 27-NOV-2001; 2001US-0333626P.

XX PA (CORI-) CORIAX CORP.

XX PI Benson DR, Kalos MD, Lodes MJ, Persing DH, Hepler WT, Jiang Y;

XX DR WPI; 2002-627435/67.

XX PT New isolated polynucleotide and pancreatic tumor polypeptides, useful for  
PT diagnosing, preventing and/or treating cancer, particularly pancreatic  
PT cancer.

XX Claim 1; SEQ ID NO 43; 300pp + Sequence Listing; English.

CC The invention relates to an isolated polynucleotide (I) comprising: (a)  
CC any of a group of over 4000 nucleotide sequences (ABV94628-ABV99145); (b)  
CC complements of (a); (c) sequences consisting of at least 20 contiguous  
CC residues of (a); (d) sequences that hybridize to (a), under moderately  
CC stringent conditions; (e) sequences having at least 75% or 90% identity  
CC to (a); or (f) degenerate variants of (a). Polypeptides (ABP68596-  
CC ABP68637) encoded by (I) and oligonucleotide can be used to detect cancer  
CC in a patient and compositions comprising polypeptides, polynucleotides,  
CC antibodies, fusion proteins, T cell populations and antigen presenting  
CC cells expressing the polypeptide are useful in treating pancreatic cancer  
CC and stimulating an immune response. The polynucleotides can be used as  
CC probes or primers for nucleic acid hybridisation, in the design and  
CC preparation of ribozyme molecules for inhibiting expression of the tumour  
CC polypeptides and proteins in the tumour cells, in vaccines and for gene  
CC therapy. Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 599 BP; 99 A; 228 C; 187 G; 85 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 1.95e-24 Length: 599  
Score: 260.00 Matches: 55  
Percent Similarity: 65.18% Conservative: 18  
Best Local Similarity: 49.11% Mismatches: 30  
Query Match: 40.69% Indels: 10  
DB: 6 Gaps: 1

US-10-657-740-1\_COPY\_51\_173 (1-123) x ABV94670 (1-599)

Qy 7 LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26  
Db 254 CTCAGCAGCGGGTCTCGAGATCCGACACTCGGACCGTGGCGGTGTCTCCCTGGAT 313  
Qy 27 VallysHisPheSerProGluAspLeuThrVallysValGlnAspAspPheValGlu 46  
Db 314 GTCAACCACTTCGCGCGCGAGCTGACGGTCAAGACCAAGGATGGCGGTGTGGAGATC 373  
Qy 47 HisGlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArg 66  
Db 374 ACCGCAAGCAGCAGGAGCGGACGAGCATGCTACATCTCTCCCGGTGCTTCAACCGG 433

```
QY 67 ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86
Db :::||| ||||| ||||| :::||| ||||| ||||| :::
434 AAATACACGCTGCCCGCGGTGTGGACCCCAAGTTCTCTCCTCCTGTCCTGAG 493
QY 87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106
Db ||| ||||| ||||| ||||| ||||| ||||| :::
494 GGCACACTGACCGTGA-GGCCCC-----CATGCCCAA 525
QY 107 ArgAlaIleProValSerArgGluGluLysProThr 118
Db ::: ||| ||||| ||||| ||||| :::
526 GCTAGCCACGAGTCCACGAGATCACCATCCAGT 561
```

Search completed: May 30, 2005, 05:14:51  
Job time : 382.402 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 30, 2005, 04:24:56 ; Search time 120.922 Seconds  
(without alignments)  
1664.391 Million cell updates/sec

Title: US-10-657-740-1\_COPY\_51\_173

Perfect score: 639

Sequence: 1 SLFRTVLDGSEIVSDRDK.....HAERAIPVSRKPTSPSS 123

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US10657740/runat\_27052005\_165254\_3343/app\_query.fasta\_1.590  
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOOEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=150 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=50  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10657740@cgn2\_1.1.219 @runat\_27052005\_165254\_3343 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=10 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	639	100.0	1114	4	US-09-949-016-5220
2	367	57.4	7775	4	US-09-949-016-16962
3	354	55.4	691	4	US-09-949-016-2545
4	354	55.4	856	4	US-09-513-999C-3891
5	354	55.4	893	4	US-09-513-999C-14942
6	354	55.4	911	4	US-09-513-999C-14938
7	354	55.4	913	4	US-09-513-999C-14937
8	354	55.4	927	4	US-09-513-999C-14936
9	354	55.4	942	4	US-09-513-999C-3890
10	330	51.6	695	4	US-09-513-999C-3892
11	309	48.4	826	4	US-09-621-976-3172
12	261.5	40.9	789	4	US-09-023-655-1300

13	261.5	40.9	845	4	US-09-949-016-4992	Sequence 4992, Ap
14	261.5	40.9	1380	3	US-09-513-783A-169	Sequence 169, App
15	260	40.7	1379	3	US-09-553-498-5	Sequence 5, Appli
16	260	40.7	1379	3	US-09-618-869-5	Sequence 5, Appli
17	188	29.4	7125	4	US-09-949-016-14287	Sequence 14287, A
18	182.5	28.6	393	4	US-09-513-999C-10503	Sequence 10503, A
19	159	24.9	1627	2	US-08-900-407-2	Sequence 2, Appli
20	158	24.7	601	4	US-09-949-016-91317	Sequence 91317, A
21	158	24.7	763	4	US-09-949-016-2623	Sequence 2623, Ap
22	158	24.7	4763	4	US-09-949-016-14365	Sequence 14365, A
23	149.5	23.4	333	4	US-09-513-999C-10502	Sequence 10502, A
24	142	22.2	601	4	US-09-949-016-89647	Sequence 89647, A
25	123.5	19.3	599	4	US-09-270-767-563	Sequence 563, App
26	123.5	19.3	599	4	US-09-270-767-15845	Sequence 15845, A
27	119	18.6	5688	4	US-09-949-016-16734	Sequence 16734, A
28	118	18.5	601	4	US-09-669-751-69	Sequence 69, Appli
29	118	18.5	601	4	US-09-949-016-89648	Sequence 89648, A
30	115	18.0	12214	4	US-09-949-016-17284	Sequence 17284, A
31	110.5	17.3	719	4	US-09-620-312D-669	Sequence 669, App
32	110	17.2	9615	4	US-09-949-016-17282	Sequence 17282, A
33	107	16.7	7125	4	US-09-949-016-14287	Sequence 14287, A
34	97	15.2	601	4	US-09-949-016-175645	Sequence 175645, A
35	90	14.1	420	4	US-09-107-532A-670	Sequence 670, App
36	88.5	13.8	652	1	US-07-748-761-2	Sequence 2, Appli
37	88	13.8	168971	4	US-09-949-016-13807	Sequence 13807, A
38	86	13.5	489	4	US-09-902-540-6436	Sequence 6436, Ap
39	86	13.5	3626	4	US-09-902-540-444	Sequence 444, App
40	85	13.3	576	4	US-09-489-039A-3463	Sequence 3463, Ap
41	79.5	12.4	1350	4	US-09-583-110-2262	Sequence 2262, Ap
42	79.5	12.4	1353	4	US-09-107-433-2322	Sequence 2322, Ap
43	77.5	12.1	49487	4	US-09-949-016-11770	Sequence 11770, A
44	77.5	12.1	57751	4	US-09-949-016-13631	Sequence 13631, A
45	75.5	11.8	966	4	US-09-969-532-23	Sequence 23, Appli
46	75.5	11.8	999	4	US-09-969-532-21	Sequence 21, Appli
47	75.5	11.8	1008	4	US-09-969-532-19	Sequence 19, Appli
48	75.5	11.8	1041	4	US-09-969-532-17	Sequence 17, Appli
49	75.5	11.8	1659	4	US-09-969-532-7	Sequence 7, Appli
50	75.5	11.8	1692	4	US-09-969-532-5	Sequence 5, Appli
51	75.5	11.8	1701	4	US-09-969-532-3	Sequence 3, Appli
52	75.5	11.8	1734	4	US-09-969-532-1	Sequence 1, Appli
53	75.5	11.8	1968	4	US-09-969-532-31	Sequence 31, Appli
54	75.5	11.8	2001	4	US-09-969-532-29	Sequence 29, Appli
55	75.5	11.8	2010	4	US-09-969-532-27	Sequence 27, Appli
56	75.5	11.8	2043	4	US-09-969-532-25	Sequence 25, Appli
57	75.5	11.8	2694	4	US-09-969-532-15	Sequence 15, Appli
58	75.5	11.8	2661	4	US-09-969-532-13	Sequence 13, Appli
59	75.5	11.8	2703	4	US-09-969-532-11	Sequence 11, Appli
60	75.5	11.8	2736	4	US-09-969-532-9	Sequence 9, Appli
61	75.5	11.8	3411	4	US-09-969-532-33	Sequence 33, Appli
62	72.5	11.3	3402	1	US-08-480-917-1	Sequence 1, Appli
63	72.5	11.3	3402	1	US-09-138-736-1	Sequence 1, Appli
64	72.5	11.3	3402	3	US-08-988-242-1	Sequence 1, Appli
65	72.5	11.3	139336	4	US-09-949-016-11782	Sequence 11782, A
66	72.5	11.3	139952	4	US-09-949-016-13280	Sequence 13280, A
67	71.5	11.2	465	4	US-09-252-991A-12236	Sequence 12236, A
68	71.5	11.2	495	4	US-09-252-991A-12170	Sequence 12170, A
69	71.5	11.2	522	4	US-09-252-991A-12373	Sequence 12373, A
70	71.5	11.2	1354	3	US-08-456-2008-4	Sequence 4, Appli
71	71.5	11.2	3563	3	US-09-041-886-20	Sequence 20, Appli
72	71.5	11.2	3596	3	US-08-779-801-5	Sequence 5, Appli
73	71.5	11.2	3596	3	US-09-298-441-5	Sequence 5, Appli
74	71.5	11.2	3632	2	US-08-779-801-3	Sequence 3, Appli
75	71.5	11.2	3632	2	US-08-779-801-4	Sequence 4, Appli
76	71.5	11.2	3632	3	US-09-298-441-3	Sequence 3, Appli
77	71.5	11.2	3632	3	US-09-298-441-4	Sequence 4, Appli
78	71.5	11.2	7791	3	US-08-149-097D-23	Sequence 23, Appli
79	71.5	11.2	7791	3	US-08-949-386-23	Sequence 23, Appli
80	71.5	11.2	7791	3	US-08-450-562-23	Sequence 23, Appli
81	71.5	11.2	7791	3	US-08-984-709A-23	Sequence 23, Appli
82	71.5	11.2	7791	3	US-08-450-272-23	Sequence 23, Appli
83	71.5	11.2	7791	3	US-08-450-273-23	Sequence 23, Appli
84	71.5	11.2	7808	2	US-08-149-097D-22	Sequence 22, Appli
85	71.5	11.2	7808	3	US-08-949-386-22	Sequence 22, Appli

```
c 86 71.5 11.2 7808 3 US-08-450-562-22
c 87 71.5 11.2 7808 3 US-08-984-709A-22
c 88 71.5 11.2 7808 3 US-08-450-272-22
c 89 71.5 11.2 7808 4 US-08-450-272-22
c 90 71.5 11.2 7815 4 US-09-949-016-3629
c 91 71.5 11.2 7815 4 US-09-949-016-3630
c 92 71 11.1 1278 4 US-09-902-540-8457
c 93 71 11.1 8590 4 US-09-902-540-889
c 94 71 11.1 1830121 4 US-09-557-884-1
c 95 71 11.1 1830121 4 US-09-643-990A-1
c 96 71 11.1 4403765 3 US-09-103-840A-2
c 97 71 11.1 4411529 3 US-09-103-840A-1
c 98 70.5 11.0 585 4 US-09-248-796A-2425
c 99 70 11.0 601 4 US-09-949-016-142871
c 100 70 11.0 601 4 US-09-949-016-182120
c 101 70 11.0 265038 4 US-09-949-016-15779
c 102 69.5 10.9 42931 4 US-08-311-731A-129
c 103 69 10.8 2379 4 US-09-543-681A-2445
c 104 69 10.8 421491 4 US-09-949-016-12805
c 105 69 10.8 421494 4 US-09-949-016-14060
c 106 68.5 10.7 327 4 US-09-134-000C-1375
c 107 68.5 10.7 576 1 US-08-086-428B-34
c 108 68.5 10.7 576 2 US-08-468-570-34
c 109 68.5 10.7 576 2 US-08-290-665A-34
c 110 68.5 10.7 576 4 US-08-466-601A-34
c 111 68.5 10.7 576 5 PCT-US95-10398-34
c 112 68.5 10.7 609 4 US-09-248-796A-1874
c 113 68 10.6 531 4 US-09-252-991A-9454
c 114 68 10.6 804 4 US-09-902-540-8110
c 115 68 10.6 834 4 US-09-252-991A-9477
c 116 68 10.6 1425 4 US-09-252-991A-9353
c 117 68 10.6 1576 4 US-09-949-016-1340
c 118 68 10.6 2214 4 US-09-902-540-4124
c 119 68 10.6 2522 3 US-09-058-389A-1
c 120 68 10.6 2522 3 US-09-611-781-1
c 121 68 10.6 6890 4 US-09-902-540-828
c 122 68 10.6 23091 4 US-09-902-540-1204
c 123 68 10.6 1664976 4 US-08-916-421B-1
c 124 68 10.6 1664976 4 US-09-692-570-1
c 125 67.5 10.6 819 4 US-09-107-532A-1887
c 126 67.5 10.6 1185 4 US-09-248-796A-3762
c 127 67.5 10.6 5517 3 US-09-120-663-1
c 128 67.5 10.6 5517 3 US-09-431-614-7
c 129 67.5 10.6 6789 4 US-09-269-446D-43
c 130 67.5 10.6 9556 4 US-09-902-540-929
c 131 67.5 10.6 1194 4 US-09-252-991A-13276
c 132 67 10.5 1305 4 US-09-252-991A-12685
c 133 67 10.5 1545 4 US-09-252-991A-11637
c 134 67 10.5 5410 3 US-09-221-017B-70
c 135 67 10.5 20757 4 US-09-902-540-1189
c 136 67 10.5 145287 4 US-09-949-016-13530
c 137 67 10.5 145287 4 US-09-949-016-13531
c 138 66.5 10.4 1337 4 US-09-690-454-19
c 139 66.5 10.4 1930 1 US-07-982-112-1
c 140 66.5 10.4 54484 4 US-09-902-540-1272
c 141 66.5 10.4 1242 3 US-09-372-448A-3
c 142 65.5 10.3 876 4 US-09-252-991A-12184
c 143 65.5 10.3 882 4 US-09-328-352-710
c 144 65.5 10.3 1392 4 US-09-252-991A-12249
c 145 65.5 10.3 1407 4 US-09-252-991A-12361
c 146 65.5 10.3 3546 3 US-09-177-431-7008
c 147 65.5 10.3 42246 4 US-09-949-016-12177
c 148 65.5 10.3 59123 4 US-09-949-016-12177
c 149 65.5 10.3 70559 4 US-09-409-800B-1
c 150 65.5 10.3 70559 4
```

## ALIGNMENTS

## RESULT 1

US-09-949-016-5220

; Sequence 5220, Application US/09949016

; Patent No. 6812339

```
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5220
; LENGTH: 1114
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-5220
Alignment Scores:
Pred. No.: 1114-80 Length: 1114
Score: 639.00 Matches: 123
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0
US-10-657-740-1_COPY_51_173 (1-123) x US-09-949-016-5220 (1-1114)
QY 1 SerLeuPheAlaThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20
Db 220 TCCCTCTTCCGACCGCTGCTGACCTCCGGCATCTCTGAGTTGATCCGCGGCAAG 279
QY 21 PheValIlePheLeuAspValIleHisPheSerProGluAspLeuThrValLysValGln 40
Db 280 TTCGTCATCTCTCGATGTGAAGCACCTTCTCCCGAGGACCTCACCGTGAAGGTGCAG 339
QY 41 AspAspPheValGluIleHisGlyIleHisAsnGluValArgGlnAspAspHisGlyIle 60
Db 340 GAGCACTTTGTGAGATCCACGGAAGCACACGAGCGCGCAGGACGACCGGTACAT 399
QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80
Db 400 TCCCTGAGTTCCACGCGCTACCGCTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGT 459
QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100
Db 460 TGCTCCCTGCTGCGGATGGCATGCTGACCTTCTGTGGCCCCCAAGATCCAGATGCGCTG 519
QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120
Db 520 GATGCCACCCAGCGGAGCGGACCATCCCGTGTCCGGGAGGAGAGGCCACCTCGGCT 579
QY 121 ProSerSer 123
Db 580 CCTCGTCC 588
RESULT 2
US-09-949-016-16962
; Sequence 16962, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
```

```
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16962
; LENGTH: 7775
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16962

Alignment Scores:
Pred. No.: 9,17e-41 Length: 7775
Score: 367.00 Matches: 70
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 57.43% Indels: 0
DB: Gaps: 0

US-10-657-740-1_COPY_51_173 (1-123) x US-09-949-016-16962 (1-7775)
QY 54 GlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeuProSerAsn 73
Db 5039 CAGGACGACCGGTACATTTCCGCTGAGTTCCACCGCGCTACCGCTGCGTCCAC 5098
QY 74 ValAspGlnSerAlaLeuSerCysSerIleuSerAlaAspGlyMetLeuThrPheCysGly 93
Db 5099 GTGGACCAAGTCGGCCCTCTTGTGCTCCCTGTGCGGATGGCATGCTGACCTTCTGTGGC 5158
QY 94 ProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProValSerArg 113
Db 5159 CCCAGATCCAGACGTGGCTGTGATGCCACCCACGCGAGCGCATCCCGTGTGCGG 5218
QY 114 GluGluLysProThrSerAlaProSerSer 123
Db 5219 GAGGAGAAGCCACCTCGGCTCCCTCGTCC 5248

RESULT 3
US-09-949-016-2545
; Sequence 2545, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2545
; LENGTH: 691
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-2545

Alignment Scores:
Pred. No.: 1.62e-40 Length: 691
Score: 354.00 Matches: 66
Percent Similarity: 78.45% Conservative: 25
Best Local Similarity: 56.90% Mismatches: 19
Query Match: 55.40% Indels: 6
DB: Gaps: 2

US-10-657-740-1_COPY_51_173 (1-123) x US-09-949-016-2545 (1-691)
QY 8 AspSerGlyIleSerGluValArgSerArgAspLysPheValIlePheLeuAspVal 27
```

```
Db 209 GACACTGGACTCTCAGAGATCGCCTGGAGAGGACAGGTTCTGTCAACCTGGATGG 268
QY 28 LysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGluIleHis 47
Db 269 AAGCACTTCTCCCAAGAGAACTCAAGTTAAGGTGTTGGGAGATGTGATTGAGGTGCAT 328
QY 48 GlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArgArg 67
Db 329 CGAAACATGAAGAGCGCAGGATGAACATGTTTTCATCTCCAGGAGATTCCACAGAAA 388
QY 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerIleuSerAlaAspGly 87
Db 389 TACCGATCCAGCTGATGATAGACCTCTCACCATTACTTCATCCTGTGCATCTGATGG 448
QY 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107
Db 449 GTCTCTCACTGTGAATGGACCAAGAAACAG-----GTCTCTGGCCCTGAGCCG 496
QY 108 AlaIleProValSerArgGluLysPro-----ThrSerAlaPro 121
Db 497 ACCATTCCATCACCCTGTGAAGAGAGCTGTGTGTCCAGCGAGCCCCC 544

RESULT 4
US-09-513-999C-3891
; Sequence 3891, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 3891
; LENGTH: 856
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4..753
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 51
; OTHER INFORMATION: s=g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 117
; OTHER INFORMATION: s=g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 118
; OTHER INFORMATION: k=g or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 159
; OTHER INFORMATION: r=a or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 160
; OTHER INFORMATION: s=g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 162
; OTHER INFORMATION: r=a or g
; FEATURE:
; NAME/KEY: misc_feature
```

```
; LOCATION: 183
; OTHER INFORMATION: r=a or g
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 16
; OTHER INFORMATION: Xaa=Asp or Glu
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 38
; OTHER INFORMATION: Xaa=His or Gln
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 39
; OTHER INFORMATION: Xaa=Ala or Ser
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 53
; OTHER INFORMATION: Xaa=Glu or Gln
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 60
; OTHER INFORMATION: Xaa=Glu or Gly
US-09-513-999C-3891

Alignment Scores:
Pred. No.: 2,24e-40 Length: 856
Score: 354.00 Matches: 65
Percent Similarity: 78.45% Conservative: 25
Best Local Similarity: 56.90% Mismatches: 19
Query Match: 55.40% Indels: 6
DB: 4 Gaps: 2

US-10-657-740-1_COPY_51_173 (1-123) x US-09-513-999C-3891 (1-856)

QY 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27
Db 412 GACACTGGAGCTCTCAGAGATCGCGCTGGAGAGACAGGTTCTCTGTCACACCTGGATGTG 471
QY 28 LysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGluLeuHis 47
Db 472 AGCACTTCTCCCGAGAGAACTCAAGTTAAGTGTGGGAGATGTGATTGAGTGCAT 531
QY 48 GlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArgArg 67
Db 532 GGAATAACATGAAGAGCGCCAGCATGAACATGTTTCACTCCAGGAGTTCACAGGAAA 591
QY 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87
Db 592 TACCGATCCCGATGATGATAGACCTCTCACCATTTACTTCTCATCTGTCATCTGATGG 651
QY 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107
Db 652 GTCTCTCACTGTGAATGACCAAGAAACAG-----GTCTCTGGCCCTGAGCGC 699
QY 108 AlaIleProValSerArgGluLysPro-----ThrSerAlaPro 121
Db 700 ACCATTCCCATCACCCTGAGAGAGAGCTGTCTGTCCACCGAGCCCC 747

RESULT 5
US-09-513-999C-14942
; Sequence 14942, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 14938
; LENGTH: 911
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-513-999C-14942

Alignment Scores:
Pred. No.: 2,47e-40 Length: 911
Score: 354.00 Matches: 66
Percent Similarity: 78.45% Conservative: 25
Best Local Similarity: 56.90% Mismatches: 19
Query Match: 55.40% Indels: 6
DB: 4 Gaps: 2

; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 14942
; LENGTH: 893
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-513-999C-14942

Alignment Scores:
Pred. No.: 2,39e-40 Length: 893
Score: 354.00 Matches: 66
Percent Similarity: 78.45% Conservative: 25
Best Local Similarity: 56.90% Mismatches: 19
Query Match: 55.40% Indels: 6
DB: 4 Gaps: 2

US-10-657-740-1_COPY_51_173 (1-123) x US-09-513-999C-14942 (1-893)

QY 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27
Db 449 GACACTGGAGCTCTCAGAGATCGCGCTGGAGAGACAGGTTCTCTGTCACACCTGGATGTG 508
QY 28 LysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGluLeuHis 47
Db 509 AAGCACTTCTCCCGAGAGAACTCAAGTTAAGTGTGGGAGATGTGATTGAGTGCAT 568
QY 48 GlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArgArg 67
Db 569 GGAATAACATGAAGAGCGCCAGCATGAACATGTTTCACTCCAGGAGTTCACAGGAAA 628
QY 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87
Db 629 TACCGATCCCGATGATGATAGACCTCTCACCATTTACTTCTCATCTGTCATCTGATGG 688
QY 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107
Db 689 GTCTCTCACTGTGAATGACCAAGAAACAG-----GTCTCTGGCCCTGAGCGC 736
QY 108 AlaIleProValSerArgGluLysPro-----ThrSerAlaPro 121
Db 737 ACCATTCCCATCACCCTGAGAGAGAGCTGTCTGTCCACCGAGCCCC 784

RESULT 6
US-09-513-999C-14938
; Sequence 14938, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 14938
; LENGTH: 911
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-513-999C-14938

Alignment Scores:
Pred. No.: 2,47e-40 Length: 911
Score: 354.00 Matches: 66
Percent Similarity: 78.45% Conservative: 25
Best Local Similarity: 56.90% Mismatches: 19
Query Match: 55.40% Indels: 6
DB: 4 Gaps: 2
```

```

US-10-657-740-1_COPY_51_173 (1-123) x US-09-513-999C-14938 (1-911)
Qy 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27
Db 467 GACACTGACTCTCAGAGATCGCGCTGGAGAGGACAGGTTCTCTGTCAACCTGGATGTG 526
Qy 28 LysHisPheSerProGluAspLeuThrValIleValGlnAspPheValGluIleHis 47
Db 527 AAGCACTTCTCCCGAGAGAACTCAAAGTTAAGGTGTTGGAGATGTGATTGAGGTGCAT 586
Qy 48 GlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArgArg 67
Db 587 GGAACAACATGAAGAGCGCCAGGATGAACATGGTTTCATCTCCAGGAGTTCCACAGGAAA 646
Qy 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87
Db 647 TACCGATCCAGCTGATGATGAGACCTCTCACCATTACTTCTATCCCTGTCATCTGATGG 706
Qy 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107
Db 707 GTCTCTCACTGTGAATGGACCAAGGAAACAG-----GTCTCTGGCCCTGAGCGC 754
Qy 108 AlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 121
Db 755 ACCATTCCCATCACCCTGTGAGAGAGAGCTGTGTGTACCCGAGCCCC 802

RESULT 7
US-09-513-999C-14937
; Sequence 14937, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR FILING DATE: 2000-02-24
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 14937
; LENGTH: 913
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
; LOCATION: 203
; OTHER INFORMATION: n=a, g, c or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 204
; OTHER INFORMATION: n=a, g, c or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 330
; OTHER INFORMATION: k=g or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 372
; OTHER INFORMATION: r=a or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 390
; OTHER INFORMATION: k=g or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 394
; OTHER INFORMATION: s=g or c
US-09-513-999C-14937

US-10-657-740-1_COPY_51_173 (1-123) x US-09-513-999C-14938 (1-911)
Qy 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27
Db 467 GACACTGACTCTCAGAGATCGCGCTGGAGAGGACAGGTTCTCTGTCAACCTGGATGTG 526
Qy 28 LysHisPheSerProGluAspLeuThrValIleValGlnAspPheValGluIleHis 47
Db 527 AAGCACTTCTCCCGAGAGAACTCAAAGTTAAGGTGTTGGAGATGTGATTGAGGTGCAT 586
Qy 48 GlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArgArg 67
Db 587 GGAACAACATGAAGAGCGCCAGGATGAACATGGTTTCATCTCCAGGAGTTCCACAGGAAA 646
Qy 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87
Db 647 TACCGATCCAGCTGATGATGAGACCTCTCACCATTACTTCTATCCCTGTCATCTGATGG 706
Qy 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107
Db 707 GTCTCTCACTGTGAATGGACCAAGGAAACAG-----GTCTCTGGCCCTGAGCGC 754
Qy 108 AlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 121
Db 755 ACCATTCCCATCACCCTGTGAGAGAGAGCTGTGTGTACCCGAGCCCC 802

RESULT 8
US-09-513-999C-14936
; Sequence 14936, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR FILING DATE: 2000-02-24
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 14936
; LENGTH: 927
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-513-999C-14936

US-10-657-740-1_COPY_51_173 (1-123) x US-09-513-999C-14937 (1-913)
Qy 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27
Db 469 GACACTGGACTCTCAGAGATCGCGCTGGAGAGGACAGGTTCTCTGTCAACCTGGATGTG 528
Qy 28 LysHisPheSerProGluAspLeuThrValIleValGlnAspPheValGluIleHis 47
Db 529 AAGCACTTCTCCCGAGAGAACTCAAAGTTAAGGTGTTGGAGATGTGATTGAGGTGCAT 588
Qy 48 GlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArgArg 67
Db 589 GGAACAACATGAAGAGCGCCAGGATGAACATGGTTTCATCTCCAGGAGTTCCACAGGAAA 648
Qy 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87
Db 649 TACCGATCCAGCTGATGATGAGACCTCTCACCATTACTTCTATCCCTGTCATCTGATGG 708
Qy 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107
Db 709 GTCTCTCACTGTGAATGGACCAAGGAAACAG-----GTCTCTGGCCCTGAGCGC 756
Qy 108 AlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 121
Db 757 ACCATTCCCATCACCCTGTGAGAGAGAGCTGTGTGTACCCGAGCCCC 804

US-10-657-740-1_COPY_51_173 (1-123) x US-09-513-999C-14937 (1-913)
Qy 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27
Db 469 GACACTGGACTCTCAGAGATCGCGCTGGAGAGGACAGGTTCTCTGTCAACCTGGATGTG 528
Qy 28 LysHisPheSerProGluAspLeuThrValIleValGlnAspPheValGluIleHis 47
Db 529 AAGCACTTCTCCCGAGAGAACTCAAAGTTAAGGTGTTGGAGATGTGATTGAGGTGCAT 588
Qy 48 GlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArgArg 67
Db 589 GGAACAACATGAAGAGCGCCAGGATGAACATGGTTTCATCTCCAGGAGTTCCACAGGAAA 648
Qy 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87
Db 649 TACCGATCCAGCTGATGATGAGACCTCTCACCATTACTTCTATCCCTGTCATCTGATGG 708
Qy 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107
Db 709 GTCTCTCACTGTGAATGGACCAAGGAAACAG-----GTCTCTGGCCCTGAGCGC 756
Qy 108 AlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 121
Db 757 ACCATTCCCATCACCCTGTGAGAGAGAGCTGTGTGTACCCGAGCCCC 804

US-10-657-740-1_COPY_51_173 (1-123) x US-09-513-999C-14936 (1-927)
Qy 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27
Db 483 GACACTGGACTCTCAGAGATCGCGCTGGAGAGGACAGGTTCTCTGTCAACCTGGATGTG 542
Qy 28 LysHisPheSerProGluAspLeuThrValIleValGlnAspPheValGluIleHis 47

```

Db 543 AAGCACTTCTCCAGAGAACTCAAAGTTAAGGTGGAGATGGATTGAGTGCAT 602  
QY 48 GlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArg 67  
Db 603 GGAACAATGAAGAGCCAGAGATGAACATGGTTTCATCTCCAGGAGATTCCACAGAAA 662  
QY 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87  
Db 663 TACCGATCCAGCTGATGTAGACCTCTCACCATTAATCTATCTCTGTCATCTGATGG 722  
QY 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107  
Db 723 GTCTCTACTGTGAATGACCAAGAAACAG-----GTCTCTGCCCTGAGCGC 770  
QY 108 AlaIleProValSerArgGluLysPro-----ThrSerAlaPro 121  
Db 771 ACCATTCCCATCACCCTGAGAGAAAGCTTGCTGTCACCGAGCCCC 818  
RESULT 9  
US-09-513-999C-3890  
; Sequence 3890, Application US/09513999C  
; Patent No. 6783961  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Duclert, A.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
; FILE REFERENCE: 59.US2.REG  
; CURRENT APPLICATION NUMBER: US/09/513,999C  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/122,487  
; NUMBER OF SEQ ID NOS: 36681  
; SOFTWARE: Patent.pm  
; SEQ ID NO 3890  
; LENGTH: 942  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 315..839  
US-09-513-999C-3890  
Alignment Scores:  
Pred. No.: 2,59e-40 Length: 942  
Score: 354.00 Matches: 66  
Percent Similarity: 78.45% Conservative: 25  
Best Local Similarity: 56.90% Mismatches: 19  
Query Match: 55.40% Indels: 6  
DB: 4 Gaps: 2  
US-10-657-740-1\_COPY\_51\_173 (1-123) x US-09-513-999C-3890 (1-942)  
QY 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27  
Db 498 GACATCGACTCTCAGATGCGCTGGAGAGACAGGTTCTCTGTCAACTGGATGTG 557  
QY 28 LysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHis 47  
Db 558 AAGCACTTCTCCAGAGAACTCAAAGTTAAGGTGGAGATGGATTGAGTGCAT 617  
QY 48 GlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArg 67  
Db 618 GGAACAATGAAGAGCCAGAGATGAACATGGTTTCATCTCCAGGAGTTCCACAGAAA 677  
QY 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87  
Db 678 TACCGATCCAGCTGATGTAGACCTCTCACCATTAATCTATCTCTGTCATCTGATGG 737  
QY 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107  
Db 738 GTCTCTACTGTGAATGACCAAGAAACAG-----GTCTCTGCCCTGAGCGC 785

QY 108 AlaIleProValSerArgGluLysPro-----ThrSerAlaPro 121  
Db 786 ACCATTCCCATCACCCTGAGAGAAAGCTTGCTGTCACCGAGCCCC 833  
RESULT 10  
US-09-513-999C-3892  
; Sequence 3892, Application US/09513999C  
; Patent No. 6783961  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Duclert, A.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
; FILE REFERENCE: 59.US2.REG  
; CURRENT APPLICATION NUMBER: US/09/513,999C  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/122,487  
; NUMBER OF SEQ ID NOS: 36681  
; SOFTWARE: Patent.pm  
; SEQ ID NO 3892  
; LENGTH: 695  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 269..592  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 44  
; OTHER INFORMATION: s=g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 45  
; OTHER INFORMATION: s=g or c  
US-09-513-999C-3892  
Alignment Scores:  
Pred. No.: 4.19e-37 Length: 695  
Score: 330.00 Matches: 62  
Percent Similarity: 77.27% Conservative: 23  
Best Local Similarity: 56.36% Mismatches: 19  
Query Match: 51.64% Indels: 6  
DB: 4 Gaps: 2  
US-10-657-740-1\_COPY\_51\_173 (1-123) x US-09-513-999C-3892 (1-695)  
QY 14 ValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPheSerProGlu 33  
Db 269 ATCGGCTGGAGAGAGACAGGTTCTCTGTCAACCTGGATGTGAGTGCATGGAACATGAGAGCGC 328  
QY 34 AspLeuThrValLysValGlnAspAspPheValGluIleHisGlyLysHisAsnGluArg 53  
Db 329 GAACTCAAAGTTAAGGTGGAGATGGATTGAGTGCATGGAACATGAGAGCGC 388  
QY 54 GlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeuProSerAsn 73  
Db 389 CAGGATGAACATGGTTTCATCTCCAGGAGATTCCACAGGAAATACCGGATCCAGCTGAT 448  
QY 74 ValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGly 93  
Db 449 GTAGACCTCTCACCATTAATCTATCTGATGGGTCTCTACTGTGATGGA 508  
QY 94 ProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProValSerArg 113  
Db 509 CCAAGAAACAG-----GTCTCTGCCCTGAGCGCACCATTCCCATCCCGCT 556  
QY 114 GluGluLysPro-----ThrSerAlaPro 121  
Db 557 GAAGAGAGACCTGCTGTCACCGAGCCCC 586

## RESULT 11

US-09-621-976-3172  
 ; Sequence 3172, Application US/09621976  
 ; Patent No. 6639063  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dumas Milne Edwards, J.B.  
 ; APPLICANT: Jobert, S.  
 ; APPLICANT: Giordano, J.Y.  
 ; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
 ; FILE REFERENCE: GENSET.054PR2  
 ; CURRENT APPLICATION NUMBER: US/09/621,976  
 ; CURRENT FILING DATE: 2000-07-21  
 ; NUMBER OF SEQ ID NOS: 19335  
 ; SOFTWARE: Patent.pm  
 ; SEQ ID NO 3172  
 ; LENGTH: 826  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 557..736  
 ; US-09-621-976-3172

Alignment Scores:  
 Pred. No.: 5 23e-34 Length: 826  
 Score: 309.00 Matches: 63  
 Percent Similarity: 76.99% Conservative: 24  
 Best Local Similarity: 55.75% Mismatches: 19  
 Query Match: 48.36% Indels: 7  
 DB: 4 Gaps: 1

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-09-621-976-3172 (1-826)

Qy 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27  
 Db 454 GACACTGACTCTCAGAGATCGCCTGGAGAGACAGAGTTCTCTGTCACTGGATGTG 513  
 Qy 28 LysHisPheSerProGluAspLeuThrValIleValGlnAspPhe-ValGluIleHi 47  
 Db 514 AAGCACTTCTCCCGACAGAGAACTCAAGTTAAGGTGTTGGGAGATGTCGATTGAGTGCA 573  
 Qy 47 sGlyIleHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArg 67  
 Db 574 TGGAAACATCAAGAGCCCGAGTAGAACATGGTTTCACTCCAGGGAGTCCACAGGAA 633  
 Qy 67 gTyArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspG1 87  
 Db 634 ATACCGGATCCAGCTGATGTAGACCTCTCACCATTAATTCTATCCCTGTCTATGTGATGG 693  
 Qy 87 yMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThr-HisAlaGluA 107  
 Db 694 GGTCTCACTGTGAATGACCAAGAAACAG-----GTCTCTGGGCCCTGAGC 741  
 Qy 107 ArgAlaIleProValSer-ArgGluGluLysPro 117  
 Db 742 GCACATTTCCATCATCCCGCTGAAGAGAAGCCT 774

## RESULT 12

US-09-623-655-1300  
 ; Sequence 1300, Application US/09023655  
 ; Patent No. 6607879  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cocks, Benjamin G.  
 ; APPLICANT: Susan G. Stuart  
 ; APPLICANT: Jeffrey J. Seilhamer  
 ; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE  
 ; TITLE OF INVENTION: EXPRESSION  
 ; NUMBER OF SEQUENCES: 1508  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESS: INCYTE PHARMACEUTICALS, INC.  
 ; STREET: 3174 PORTER DRIVE  
 ; CITY: PALO ALTO  
 ; STATE: CALIFORNIA

COUNTRY: USA  
 ZIP: 94304  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/023,655  
 FILING DATE: HEREWITH  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE:  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Zeller, Karen J.  
 REGISTRATION NUMBER: 37,071  
 REFERENCE/DOCKET NUMBER: PA-0001 US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (650) 855-0555  
 TELEFAX: (650) 845-4166  
 INFORMATION FOR SEQ ID NO: 1300:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 789 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 IMMEDIATE SOURCE:  
 LIBRARY: GENBANK  
 CLONE: g32477  
 US-09-023-655-1300

Alignment Scores:  
 Pred. No.: 2.72e-27 Length: 789  
 Score: 261.50 Matches: 55  
 Percent Similarity: 65.18% Conservative: 18  
 Best Local Similarity: 49.11% Mismatches: 30  
 Query Match: 40.92% Indels: 10  
 DB: 4 Gaps: 1

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-09-023-655-1300 (1-789)

Qy 7 LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26  
 Db 293 CTCAGCAGCGGGTCTCGAGATCCGCACACTCGGACCGCTGGCGGTGTCCTGGAT 352  
 Qy 27 ValIleHisPheSerProGluAspLeuThrValIleValGlnAspAspPheValGluIle 46  
 Db 353 GTCAACCACTTTCGCCCGGACGAGTGCAGCTCAAGACCAAGGATGGCGTGTGGAGATC 412  
 Qy 47 HisGlyIleHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArg 66  
 Db 413 ACCGCGACGACAGGAGCGGACGACGAGATGGTACATCTCCCGTGTCTTACGGCG 472  
 Qy 67 ArgTyArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86  
 Db 473 AATACAGCTGCCCGCGGTGTGGACCCCAAGTTTCTCTCTCTCTCTCTCTCTCTGAG 532  
 Qy 87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106  
 Db 533 GGCACACTGACCGTGA-GGCCCC-----CATGCCCAA 564  
 Qy 107 ArgAlaIleProValSerArgGluGluLysProThr 118  
 Db 565 GCTAGCAGCGAGTCCCAACGAGATCACATCCCACT 600

## RESULT 13

US-09-949-016-4992  
 ; Sequence 4992, Application US/09949016  
 ; Patent No. 6812339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VENTER, J. Craig et al.

;; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
;; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
;; FILE REFERENCE: CL001307  
;; CURRENT APPLICATION NUMBER: US/09/949,016  
;; CURRENT FILING DATE: 2000-04-14  
;; PRIOR APPLICATION NUMBER: 60/241,755  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/237,768  
;; PRIOR FILING DATE: 2000-10-03  
;; PRIOR APPLICATION NUMBER: 60/231,498  
;; PRIOR FILING DATE: 2000-09-08  
;; NUMBER OF SEQ ID NOS: 207012  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 4992  
;; LENGTH: 845  
;; TYPE: DNA  
;; ORGANISM: Human  
US-09-949-016-4992

Alignment Scores:  
Pred. No.: 3,02e-27 Length: 845  
Score: 261.50 Matches: 55  
Percent Similarity: 65.18% Conservative: 18  
Best Local Similarity: 49.11% Mismatches: 30  
Query Match: 40.92% Indels: 10  
DB: 4 Gaps: 1

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-09-949-016-4992 (1-845)

QY 7 LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26  
Db 348 CTCAGCAGCGGGTCTCGGAGATCCGGACACTGCGGACCGTGGCGGTGCTCGCTGGAT 407  
QY 27 VallysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGluLe 46  
Db 408 GTCAACCACTTCGCCCGGACGAGCTGACGCTCAGACCAAGATGGCGTGTGGAGATC 467  
QY 47 HisGlyLysHisAsnGluArgGlnAspHisGlyTyrlleSerArgGluPheHisArg 66  
Db 468 ACCGGAACGACGAGGAGCGGACGAGCAGCATGCTACATCTCCCGTGTTCACGCGG 527  
QY 67 ArgTyArgLeuProSerAsnValArgGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86  
Db 528 AAATACACGCTGCCCGCGTGGACCCCAAGTTTCCTCCCTGCTCCCTGAG 587  
QY 87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106  
Db 588 GGCACACTGACCGTGA-GGCCCC-----CATGCCCAA 619  
QY 107 ArgAlaIleProValSerArgGluGluLysProThr 118  
Db 620 GCTAGCCAGCGAGTCCACGAGATCACCATCCCACT 655

RESULT 14

US-09-513-783A-169  
; Sequence 169, Application US/09513783A  
; Patent No. 6416959  
; GENERAL INFORMATION:  
; APPLICANT: Giuliano, Kenneth A.  
; APPLICANT: Kapur, Ravi  
; TITLE OF INVENTION: A System for Cell Based Screening  
; FILE REFERENCE: 97-022-L1  
; CURRENT APPLICATION NUMBER: US/09/513,783A  
; CURRENT FILING DATE: 2000-02-25  
; NUMBER OF SEQ ID NOS: 180  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 169  
; LENGTH: 1380  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GFP-HSP27  
; NAME/KEY: CDS

;; LOCATION: (1)..(1380)  
US-09-513-783A-169

Alignment Scores:  
Pred. No.: 6.37e-27 Length: 1380  
Score: 261.50 Matches: 55  
Percent Similarity: 65.18% Conservative: 18  
Best Local Similarity: 49.11% Mismatches: 30  
Query Match: 40.92% Indels: 10  
DB: 3 Gaps: 1

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-09-513-783A-169 (1-1380)

QY 7 LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26  
Db 1003 CTCAGCAGCGGGTCTCGGAGATCCGGACACTGCGGACCGTGGCGGTGCTCGCTGGAT 1062  
QY 27 VallysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGluLe 46  
Db 1063 GTCAACCACTTCGCCCGGACGAGCTGACGCTCAGACCAAGATGGCGTGTGGAGATC 1122  
QY 47 HisGlyLysHisAsnGluArgGlnAspHisGlyTyrlleSerArgGluPheHisArg 66  
Db 1123 ACCGGAACGACGAGGAGCGGACGAGCATGCTACATCTCCCGTGTTCACGCGG 1182  
QY 67 ArgTyArgLeuProSerAsnValArgGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86  
Db 1183 AAATACACGCTGCCCGCGTGGACCCCAAGTTTCCTCCCTGCTCCCTGAG 1242  
QY 87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106  
Db 1243 GGCACACTGACCGTGA-GGCCCC-----CATGCCCAA 1274  
QY 107 ArgAlaIleProValSerArgGluGluLysProThr 118  
Db 1275 GCTAGCCAGCGAGTCCACGAGATCACCATCCCACT 1310

RESULT 15

US-09-553-498-5  
; Sequence 5, Application US/09553498  
; Patent No. 6309861  
; GENERAL INFORMATION:  
; APPLICANT: Ambrosius, Dorothee  
; APPLICANT: Rudolph, Rainer  
; APPLICANT: Schasffner, Joerg  
; APPLICANT: Schwarz, Elisabeth  
; TITLE OF INVENTION: Process for the production of naturally folded and secreted prote  
; FILE REFERENCE: Case 20379  
; CURRENT APPLICATION NUMBER: US/09/553,498  
; CURRENT FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: EP99107412.1  
; PRIOR FILING DATE: 1999-04-26  
; NUMBER OF SEQ ID NOS: 10  
; SEQ ID NO 5  
; LENGTH: 1379  
; TYPE: DNA  
; ORGANISM: E. coli  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (392)...(1090)  
US-09-553-498-5

Alignment Scores:  
Pred. No.: 1.04e-26 Length: 1379  
Score: 260.00 Matches: 57  
Percent Similarity: 65.18% Conservative: 16  
Best Local Similarity: 50.89% Mismatches: 33  
Query Match: 40.69% Indels: 6  
DB: 3 Gaps: 2

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-09-553-498-5 (1-1379)

QY 7 LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26



Db 713 CTCAGCAGCGGGTCTCGGAGATCCGACAGCGGTGATCGTGGCGGTGCTCCCTGGAC 772  
Qy 27 VallysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGluile 46  
Db 773 GTCAACCACTTCGCTCCGAGGAGCTCACAGTGAAGACCAAGAGCGGTGGAGATC 832  
Qy 47 HisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArg 66  
Db 833 ACTGCAACGACGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 892  
Qy 67 ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86  
Db 893 AAATACACGCTCCCTCCAGGTGTGACCCACCTAGTGTCTCTCCCTATCCCTGAG 952  
Qy 87 GlyMetLeuThrPheCysGly-----ProLysIleGlnThrGlyLeuAspAlaThrHis 104  
Db 953 GGCACACTTACCGTGGAGGCTCGGTGTCGCAAGACGAGTCACG-----CAGTCA 1000  
Qy 105 AlaGluArgAlaIleProValSerArgGluGluLys 116  
Db 1001 GCGGAGATCACCATTCCGGTTACTTTTCGAGGCCCGC 1036

## RESULT 16

US-09-618-869-5  
; Sequence 5, Application US/09618869  
; Patent No. 6455279  
; GENERAL INFORMATION:  
; APPLICANT: Ambrosius, Dorthie  
; APPLICANT: Rudolph, Rainer  
; APPLICANT: Schaeffner, Joerg  
; APPLICANT: Schwarz, Elisabeth  
; TITLE OF INVENTION: PROCESS FOR THE PRODUCTION OF NATURALLY FOLDED AND  
; TITLE OF INVENTION: SECRETED PROTEINS BY CO-SECRETION OF MOLECULAR  
; TITLE OF INVENTION: CHAPERONES  
; FILE REFERENCE: 20381  
; CURRENT APPLICATION NUMBER: US/09/618,869  
; CURRENT FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: EP99114811.5  
; PRIOR FILING DATE: 1999-07-29  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 1379  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (392)...(1090)  
US-09-618-869-5

Alignment Scores:  
Pred. No.: 1,04e-26 Length: 1379  
Score: 260.00 Matches: 57  
Percent Similarity: 65.18% Conservative: 16  
Best Local Similarity: 50.89% Mismatches: 33  
Query Match: 40.69% Indels: 2  
Gaps: 2

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-09-618-869-5 (1-1379)

Qy 7 LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26  
Db 713 CTCAGCAGCGGGTCTCGGAGATCCGACAGCGGTGATCGTGGCGGTGCTCCCTGGAC 772  
Qy 27 VallysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGluile 46  
Db 773 GTCAACCACTTCGCTCCGAGGAGCTCACAGTGAAGACCAAGAGCGGTGGAGATC 832  
Qy 47 HisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArg 66  
Db 833 ACTGCAACGACGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 892

Qy 67 ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86  
Db 893 AAATACACGCTCCCTCCAGGTGTGACCCACCTAGTGTCTCTCCCTATCCCTGAG 952  
Qy 87 GlyMetLeuThrPheCysGly-----ProLysIleGlnThrGlyLeuAspAlaThrHis 104  
Db 953 GGCACACTTACCGTGGAGGCTCGGTGTCGCAAGACGAGTCACG-----CAGTCA 1000  
Qy 105 AlaGluArgAlaIleProValSerArgGluGluLys 116  
Db 1001 GCGGAGATCACCATTCCGGTTACTTTTCGAGGCCCGC 1036

## RESULT 17

US-09-949-016-14287  
; Sequence 14287, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14287  
; LENGTH: 7125  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(7125)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-14287

Alignment Scores:  
Pred. No.: 2.13e-15 Length: 7125  
Score: 188.00 Matches: 37  
Percent Similarity: 75.36% Conservative: 15  
Best Local Similarity: 53.62% Mismatches: 11  
Query Match: 29.42% Indels: 6  
Gaps: 2

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-09-949-016-14287 (1-7125)

Qy 55 AspAspHisGlyTyrIleSerArgGluPheHisArgTyrArgLeuProSerAsnVal 74  
Db 4783 GATGAACATGTTTCATCTCCAGGAGTTCACAGAAATACCGATCCACCTGATGTA 4842  
Qy 75 AspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyPro 94  
Db 4843 GACCCTCTCACCATTACTTCTCCCTGTCATCTGATGGGTCTCTCACTGTGAATGACCA 4902  
Qy 95 LysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProValSerArgGlu 114  
Db 4903 AGGAACACAG-----GTCTCTGGCCCTGAGCGCAGCCATTCCTCATCCCGTGAA 4950  
Qy 115 GluLysPro-----ThrSerAlaPro 121  
Db 4951 GAGAAGCTGTGTCTCAGCGACGCCCC 4977

## RESULT 18

US-09-513-999C-10503  
; Sequence 10503, Application US/09513999C  
; Patent No. 6783961  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.

; APPLICANT: Duclert, A.  
 ; APPLICANT: Giordano, J.Y.  
 ; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins

```

/ Patent NO. 6783961
/ FILE REFERENCE: 59.US2.REG
/ CURRENT APPLICATION NUMBER: US/09/513,999C
/ CURRENT FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: US 60/122,487
/ PRIOR FILING DATE: 1999-02-26
/ NUMBER OF SEQ ID NOS: 36681
/ SOFTWARE: Patent.pm
/ SEQ ID NO 10503

```

```

/ LENGTH: 393
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 75
/ OTHER INFORMATION: s=g or c
US-09-513-999C-10503

```

Alignment Scores:	
Pred. No.:	1.57e-16
Score:	182.50
Pred. Percent Similarity:	60.23%
Best Local Similarity:	47.73%
Query Match:	28.56%
DB:	4
Length:	393
Matches:	42
Conservative:	11
Mismatches:	26
Indels:	10
Gaps:	1

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-09-513-999C-10503 (1-393)

31	SerProGluAspLeuThrValIysValGlnAspPheValGluIleHisGlyLysHis	50
32	..	
33	..	
34	..	
35	..	
36	AGCATGAACGAGCTCACGCTCAAGACCAAGATCGCTGTGGAGATCACCGGCAAGCAC	95
37	..	
38	..	
39	..	
40	..	
41	..	
42	..	
43	..	
44	..	
45	..	
46	..	
47	..	
48	..	
49	..	
50	..	
51	AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeu	70
52	..	
53	..	
54	..	
55	..	
56	..	
57	..	
58	..	
59	..	
60	..	
61	..	
62	..	
63	..	
64	..	
65	..	
66	..	
67	..	
68	..	
69	..	
70	..	
71	..	
72	..	
73	..	
74	..	
75	..	
76	..	
77	..	
78	..	
79	..	
80	..	
81	..	
82	..	
83	..	
84	..	
85	..	
86	..	
87	..	
88	..	
89	..	
90	..	
91	..	
92	..	
93	..	
94	..	
95	..	
96	GAGGAGCGCGAGGACGAGCATGGGTACATCTCCGGTGTTCACGGGAATACACGCTG	155
97	..	
98	..	
99	..	
100	..	

71	ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr	90
QY		
Db		
156	CCCCCGGTGTGGACCCCAAGTTCCTTCCTGCTCCCTGTAGGCGCACTGACC	215
QY		
91	PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro	110
QY		
216	GTGGA-GGCCCC-----CATGCCCAAGTACCCACGCA	247
Db		

Oy	111	ValSerArgGluGluLysProThr	118
		:::    :::	
Db	248	GTCCAACGAGATCACCATCCCA	271

RESULT 19  
US-08-900-407-2

Sequence 2, Application US/08900407  
Patent No. 5962262

GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer  
APPLICANT: Lal, Preeti  
APPLICANT: Corley, Neil C.  
APPLICANT: Shah, Purvi  
TITLE OF INVENTION: NEW HUMAN HEAT SHOCK 27 LIKE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Dr.

CITY: Palo Alto  
STATE: CA  
COUNTRY: USA

ZIP: 94304  
COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible

```

? OPERATING SYSTEM: DOS
? SOFTWARE: Fast-SEQ for Windows Version 2.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/900,407
? FILING DATE: Filed Herewith
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER:

```

; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0351 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415 855 8555

TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 2.

```

; SEQUENCE CHARACTERISTICS:
;     LENGTH: 1627 base pairs

```

```

;
; TYPE: nucleic acid
; STRANDEDNESS: single

```

TOPLOGY: linear  
ORIGINAL SOURCE:

INDIVIDUAL ISOLATE: I  
IMMEDIATE SOURCE:  
LIBRARY. 2222

LIBRARY: fff  
CLONE: 1362715  
US-08-900-407-2

**Alignment Scores:**

Pred. No.:	2.96e-12	Length
Score:	159.00	Match

Percent Similarity:	64.86%
Best Local Similarity:	37.84%

```
Query Match: 24.88%
DB:         2
Index
Gaps
```

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-08:

Qy 23 IlePheLeuAspValLysHisPheSerPro

D**b** 808 GTGTGTGAATGTGCACAGCTTCAAGCCA

**Qy**      **43** PheValGluIleHisGlyLysHisAsnGlu  
         ::::| | | | : | | | | |  
**np**      TACCTGGCGCCTGTATGAGTCCGCATTTC

888 TACGTTGGAGGTCTCTGCATAAACAATGAAGA  
63 GlnPheHisArgValThrValLeuProGlu

873 C G A A C C A A G A A A A T C C A G C T T C C T G C A  
 ||| :  
 928 A A C T T C A C A A G A A A A T C C A G C T T C C T G C A

Qy 83 LeuSerAlaAspGlyMetLeuThrPheCysE

Db  
988 CTTTCCCCAGAGGGTCTGCTGATCATCGAA

RESULT 20  
US-09-949-015-01217

US 09-949-018-91317  
; Sequence 91317, Application US/09949016  
: Patent No. 6812339

APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOW  
; TITLE OF INVENTION: WITH HUMAN DISEASE,

```

; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016

```

; CURRENT FILING DATE: 2000-04-14  
 ; PRIOR APPLICATION NUMBER: 60/241,755  
 ; PRIOR FILING DATE: 2000-10-20

; PRIORITY DATE: 2000-10-20  
 ; PRIOR APPLICATION NUMBER: 60/237,768  
 ; PRIOR FILING DATE: 2000-10-03

;; PRIOR APPLICATION NUMBER: 60/231,498  
;; PRIOR FILING DATE: 2000-09-08

```

; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0

```

```
; SEQ ID NO 91317
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-91317

Alignment Scores:
Pred. No.: 9.03e-13 Length: 601
Score: 158.00 Matches: 31
Percent Similarity: 60.00% Conservative: 14
Best Local Similarity: 41.33% Mismatches: 30
Query Match: 24.73% Indels: 0
Gaps: 4

US-10-657-740-1_COPY_51_173 (1-123) x US-09-949-016-91317 (1-601)

Qy 15 ArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPheSerProGluAsp 34
Db 212 CGAGNAGCAATCCACCTTCAGATCTCTGTCGACGGTGGTCCAGTTCCTCCCTGAAGAC 271
Qy 35 LeuThrValLysValGlnAspPheValGluIleHisGlyLysHisAsnGluArgGln 54
Db 272 ATCATTCATTCAGACCTTCGAAGGCTGGCTGCTGATAAAAGCAACACGGAACCAAGATG 331
Qy 55 AspAspHisGlyTyrIleSerArgGluPheHisArgTyrArgLeuProSerAsnVal 74
Db 332 GATGAGCAGCGTGTATCTCAAGAAGCTTCACCCGACAGTACAACTACCAAGATGGCGTG 391
Qy 75 AspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeu 89
Db 392 GAAATCAAGATTGTCTGCAGTCTCTGTCATGATGAATTTTG 436

RESULT 21
US-09-949-016-2623
; Sequence 2623, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2623
; LENGTH: 763
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-2623

Alignment Scores:
Pred. No.: 1.3e-12 Length: 763
Score: 158.00 Matches: 31
Percent Similarity: 60.00% Conservative: 14
Best Local Similarity: 41.33% Mismatches: 30
Query Match: 24.73% Indels: 0
Gaps: 4

US-10-657-740-1_COPY_51_173 (1-123) x US-09-949-016-2623 (1-763)

Qy 15 ArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPheSerProGluAsp 34
Db 368 CGAGNAGCAATCCACCTTCAGATCTCTGTCGACGGTGGTCCAGTTCCTCCCTGAAGAC 427
Qy 35 LeuThrValLysValGlnAspPheValGluIleHisGlyLysHisAsnGluArgGln 54
Db 428 ATCATTCATTCAGACCTTCGAAGGCTGGCTGCTGATAAAAGCAACACGGAACCAAGATG 487
Qy 55 AspAspHisGlyTyrIleSerArgGluPheHisArgTyrArgLeuProSerAsnVal 74
Db 488 GATGAGCAGCGTGTATCTCAAGAAGCTTCACCCGACAGTACAACTACCAAGATGGCGTG 547
Qy 75 AspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeu 89
Db 548 GAAATCAAGATTGTCTGCAGTCTCTGTCATGATGAATTTTG 592

RESULT 22
US-09-949-016-14365
; Sequence 14365, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14365
; LENGTH: 4763
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14365

Alignment Scores:
Pred. No.: 2.11e-11 Length: 4763
Score: 158.00 Matches: 31
Percent Similarity: 60.00% Conservative: 14
Best Local Similarity: 41.33% Mismatches: 30
Query Match: 24.73% Indels: 0
Gaps: 4

US-10-657-740-1_COPY_51_173 (1-123) x US-09-949-016-14365 (1-4763)

Qy 15 ArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPheSerProGluAsp 34
Db 2368 CGAGNAGCAATCCACCTTCAGATCTCTGTCGACGGTGGTCCAGTTCCTCCCTGAAGAC 2427
Qy 35 LeuThrValLysValGlnAspPheValGluIleHisGlyLysHisAsnGluArgGln 54
Db 2428 ATCATTCATTCAGACCTTCGAAGGCTGGCTGCTGATAAAAGCAACACGGAACCAAGATG 2487
Qy 55 AspAspHisGlyTyrIleSerArgGluPheHisArgTyrArgLeuProSerAsnVal 74
Db 2488 GATGAGCAGCGTGTATCTCAAGAAGCTTCACCCGACAGTACAACTACCAAGATGGCGTG 2547
Qy 75 AspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeu 89
Db 2548 GAAATCAAGATTGTCTGCAGTCTCTGTCATGATGAATTTTG 2592

RESULT 23
US-09-513-999C-10502
; Sequence 10502, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513.999C
```



```

; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-15845

Alignment Scores:
Pred. No.: 7.14e-08 Length: 599
Score: 123.50 Matches: 28
Percent Similarity: 52.94% Conservative: 17
Best Local Similarity: 32.94% Mismatches: 39
Query Match: 19.33% Indels: 1
DB: 4 Gaps: 1

US-10-657-740-1_COPY_51_173 (1-123) x US-09-270-767-15845 (1-599)

QY 11 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe 30
Db 260 CTAAGTCGGGTAAATTCGCGAAGGATGTTTCGAGGCAATGTGATGTCACCTGTTTC 319
QY 31 SerProGluAspLeuThrValLysValGlnAspAspPheValGluLeuHisGlyLysHis 50
Db 320 AAGCCCTATGAGATTAGCGTGAAGACCTCAGCGCACACTGTGTCGTGGAGGCCAAGCAC 379
QY 51 AsnGluArgGlnAspAspHisGlyTyrlleSerArgGluPheHisArgArgTyrArgLeu 70
Db 380 GAGAACGACGATGATGTCACACCTTCGTGGTGGCCACATCGTCAAGCGGTTCGTCCTG 439
QY 71 ProSerAsnValaspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 90
Db 440 CCCCAGGAGTACTATCCCAACGATGTCGATCGGAACCTGTCGTCGATGCCATCTTACC 499
QY 91 Phe---CysGlyPro 94
Db 500 GTCAAGTGTCCCGC 514

RESULT 27
US-09-949-016-16734
; Sequence 16734, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16734
; LENGTH: 5688
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16734

Alignment Scores:
Pred. No.: 9.56e-06 Length: 5688
Score: 119.00 Matches: 22
Percent Similarity: 74.42% Conservative: 10
Best Local Similarity: 51.16% Mismatches: 11
Query Match: 18.62% Indels: 0
DB: 4 Gaps: 0

US-10-657-740-1_COPY_51_173 (1-123) x US-09-949-016-16734 (1-5688)

QY 7 LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26
Db 2348 CTCAGCAGCGGGGTCTCGAGATCCGCGCACACTGCGGACCGCTGCGCGTGTCCCTGGAT 2407

QY 27 ValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGluLeu 46
Db 2408 GTCAACCACTTCGCCCCGCGAGCTGACGGTCAGACCAAGGATGCGGTGGTGGAGATC 2467

QY 47 HisGlyLys 49
Db 2468 ACCGGTGAG 2476

RESULT 28
US-09-669-751-69
; Sequence 69, Application US/09669751
; Patent No. 6551575
; GENERAL INFORMATION:
; APPLICANT: Greenspan, Ralph J.
; TITLE OF INVENTION: Methods for Identifying Compounds for
; TITLE OF INVENTION: Motion Sickness, Vertigo and Other Disorders Related to
; TITLE OF INVENTION: Balance and the Perception of Gravity
; FILE REFERENCE: P-NI 3864
; CURRENT APPLICATION NUMBER: US/09/669,751
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US 60/168,579
; PRIOR FILING DATE: 1999-12-02
; NUMBER OF SEQ ID NOS: 261
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 69
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Drosophila
US-09-669-751-69

Alignment Scores:
Pred. No.: 3.23e-07 Length: 495
Score: 118.00 Matches: 21
Percent Similarity: 68.29% Conservative: 7
Best Local Similarity: 51.22% Mismatches: 13
Query Match: 18.47% Indels: 0
DB: 4 Gaps: 0

US-10-657-740-1_COPY_51_173 (1-123) x US-09-669-751-69 (1-495)

QY 18 ArgAspLysPheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrVal 37
Db 368 AAGATGGCTTCACGTGGCATGGATGTGTCAGTTCAGCCCAACGATGACCGTC 427
QY 38 LysValGlnAspAspPheValGluLeuHisGlyLysHisAsnGluArgGlnAspAspHis 57
Db 428 AAGTGGTGGACACACACCTGCTGTAGAGGGCCAGCACGAGGAGCGGAGCGCCAT 487
QY 58 Gly 58
Db 488 GGA 490

RESULT 29
US-09-949-016-89648/c
; Sequence 89648, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 89648
; LENGTH: 601
```

```
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-89648

Alignment Scores:
Pred. No.: 4.33e-07 Length: 601
Score: 118.00 Matches: 21
Percent Similarity: 82.76% Conservative: 3
Best Local Similarity: 72.41% Mismatches: 5
Query Match: 18.47% Indels: 0
DB: 4 Gaps: 0

US-10-657-740-1_COPY_51_173 (1-123) x US-09-949-016-89648 (1-601)

Qy 26 AspValIleHisPheSerProGluAspLeuThrValIleValGlnAspPheValGlu 45
Db 601 GATGTGAAGCACTTCTCCCGAGGAACTCAAGATTAAAGTTGGAGATGTGATTGAG 542
Qy 46 IleHisGlyIleHisAsnGluArgGln 54
Db 541 GTGCATGGAACATGAGAGGCCAG 515

RESULT 30
US-09-949-016-17284/c
; Sequence 17284, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17284
; LENGTH: 12214
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17284

Alignment Scores:
Pred. No.: 0.000113 Length: 12214
Score: 115.00 Matches: 35
Percent Similarity: 51.35% Conservative: 22
Best Local Similarity: 31.53% Mismatches: 36
Query Match: 18.00% Indels: 18
DB: 4 Gaps: 5

US-10-657-740-1_COPY_51_173 (1-123) x US-09-949-016-17284 (1-12214)

Qy 4 ArgThrValIleAspSerGlyIleSerGluValArgSerAsp-----ArgAspIlePhe 21
Db 376 CGGCTGCTCAGGACAGTCCAGCGCTCAGGAGGACATGACATGCCAGAGCGGTTTC 317
Qy 22 ValIlePheLeuAspValIleHisPheSerProGluAspLeuThrValIleValGlnAsp 41
Db 316 CAATGAGCTGGATGCCACCGCTTCGCCCGGAGGAACCTGGTGGTGCAGGTGATGCC 257
Qy 42 AspPheValGluIleHisGlyIleHisAsn-----GluArgGlnAsp 55
Db 256 CAATGGCTGATGTCACCGGACAGCAGCAACTGACGTGAGGACCCCGAAAGGGTCAGT 197
Qy 56 AspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeuProSerAsnValAsp 75
Db 196 TACCCG-----ATGTACACAGAGGTGACCGGAAA---ATGCTACCGTCCAACCTGAGT 146
```

```
Qy 76 GlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGly----- 93
Db 145 CCTACCGCCATGACCTGCTGCTGACCCCTCGGCGACGCTGTGGGTTCAGAGCCAGTGT 86
Qy 94 -----ProIleGlnThrGly 99
Db 85 GTGGCGCTGCCCTCCTCGAAGCCCAACAGGA 53

RESULT 31
US-09-620-312D-669
; Sequence 669, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; FILE REFERENCE: Polypeptides
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 669
; LENGTH: 719
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (81)..(578)
US-09-620-312D-669

Alignment Scores:
Pred. No.: 6.62e-06 Length: 719
Score: 110.50 Matches: 23
Percent Similarity: 54.32% Conservative: 21
Best Local Similarity: 28.40% Mismatches: 34
Query Match: 17.29% Indels: 3
DB: 4 Gaps: 1

US-10-657-740-1_COPY_51_173 (1-123) x US-09-620-312D-669 (1-719)

Qy 10 GlyIleSerGluValArgSerAspAspIlePheValIlePheLeuAspValIleHis 29
Db 276 GGGGCGAGCAACATCAAGACCTAGGACGCTATGATGTTGGGTGGACGTGAGAGAC 335
Qy 30 PheSerProGluAspLeuThrValIleValGlnAspPheValGluIleHisGlyIle 49
Db 336 TTCTCCTCAGAGACATCATTTGTCACCACTCCCAACCAACCATCGAGGTGCGGCT--- 392
Qy 50 HisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArg 69
Db 393 -----GAGAAGCTGGCGGCTGACGGCACTGTCATGAACACCTTCGCTCAAGTGCAG 446
Qy 70 LeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeu 89
```

```
Db 447 CTGCGGAGGAGTGGACCCGACGTCGGTCTCTCGGGAGAGCAGCGACGCTC 506
Qy 90 Thr 90
Db 507 ACT 509

RESULT 32
US-09-949-016-17282/c
; Sequence 17282, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17282
; LENGTH: 9615
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17282

Alignment Scores:
Pred. No.: 0.000403 Length: 9615
Score: 110.00 Matches: 31
Percent Similarity: 51.58% Conservative: 18
Best Local Similarity: 32.63% Mismatches: 30
Query Match: 17.21% Indels: 16
DB: 4 Gaps: 4

US-10-657-740-1_COPY_51_173 (1-123) x US-09-949-016-17282 (1-9615)

Qy 18 ArgAspLysPheValIlePheLeuAspValLysPheSerProGluAspLeuThrVal 37
Db 9588 AGACACGGTTTCCAAATGAAGCTGGATGCCACCGCTTCGCCCGGAGGAACTGGTGTG 9529
Qy 38 LysValGlnAspAspPheValGluLeuHisGlyLysHisAsn----- 51
Db 9528 CAGGTGGATGCCAATGCTGTGACCGACAGCAGCACTGGAGCTCAGGACCCG 9469
Qy 52 GluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeuPro 71
Db 9468 GAAAGGGTCAGTTACCGC-----ATGTCACAGAGGTGCACCGAAA---ATGCTACCG 9418
Qy 72 SerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPhe 91
Db 9417 TCCAACCTGAGTCTCTACCGCATGACCTGCTGCTGACCCCTCCGGCGCAGCTGTGGTGC 9358
Qy 92 CysGly-----ProLysIleGlnThrGly 99
Db 9357 AGAGGCCAGTGTGTGGCGCTGGCCCTCTCCCTGAAGCCCAACAGGA 9313

RESULT 33
US-09-949-016-14287/c
; Sequence 14287, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
```

```
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14287
; LENGTH: 7125
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(7125)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14287

Alignment Scores:
Pred. No.: 0.000682 Length: 7125
Score: 107.00 Matches: 18
Percent Similarity: 69.05% Conservative: 11
Best Local Similarity: 42.86% Mismatches: 13
Query Match: 16.74% Indels: 0
DB: 4 Gaps: 0

US-10-657-740-1_COPY_51_173 (1-123) x US-09-949-016-14287 (1-7125)

Qy 9 SerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLys 28
Db 130 GCAGGGCCCTCCGAGCTTAGCTCAGTGAGGGCAGTTCCAGGCATTTCTGGATGTGAGC 71
Qy 29 HisPheSerProGluAspLeuThrValLysValGlnAspPheValGluLeuHisGly 48
Db 70 CACTTTATCCCCAGCAGAGTGACTGTGAGGACTGTGGATAACCTGCTGGAGGTGTCTGCC 11
Qy 49 LysHis 50
Db 10 CGGCAC 5

RESULT 34
US-09-949-016-175645
; Sequence 175645, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 175645
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-175645

Alignment Scores:
Pred. No.: 0.000417 Length: 601
Score: 97.00 Matches: 29
Percent Similarity: 39.33% Conservative: 6
Best Local Similarity: 32.58% Mismatches: 13
Query Match: 15.18% Indels: 41
DB: 4 Gaps: 2
```

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-09-949-016-175645 (1-601)

QY 40 GlnAspPhe---ValGluIleHisGlyIysHisAsnGluArgGlnAspAspHisGly 58  
 |||||  
 Db 333 CAGTCTGATTTCCTCTCCCTCCCAAGCGACACGAGGACCGCGAGGACGAGCATGGC 392  
 |||||  
 QY 59 TyrIleSerArgGluPheHisArgArgTyr-----68  
 |||||  
 Db 393 TACATCTCCCGGTGCTTCACGCGGAATACACGTGATCTCTGGCGCCAGGTCGGGGTGGG 452  
 |||||  
 QY 68 -----68  
 Db 453 TGGGTGGCGTGGGGTGGGGTCAGGGAACAGGCGCACAGGACCCACCCCGGTGTGTAATGT 512  
 |||||  
 QY 69 -----ArgLeuProSerAsnValAspGlnSerAlaLeu 79  
 |||||  
 Db 513 AACGCTTGCCITTTCTCTCTGCACTGCAGGCTGCCCGCCGGGTGTGGACCCACCCCAAGT 572  
 |||||  
 QY 79 uSerCysSerLeuSerAlaAspGly 87  
 |||||  
 Db 573 TTCTCTCTCCCTGCTCCCTGAGGCG 597

RESULT 35

US-09-107-532A-670  
; Sequence 670, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD/ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-012  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 670:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 420 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: circular  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Enterococcus faecium  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (B) LOCATION 1...420  
; SEQUENCE DESCRIPTION: SEQ ID NO: 670:



## TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 861-3000  
 TELEFAX: (202) 822-0944  
 TELEX: 6714627 CUSH  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 652 base pairs  
 TYPE: NUCLEIC ACID  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 US-07-748-761-2

## Alignment Scores:

Pred. No.: 0.0076 Length: 652  
 Score: 88.50 Matches: 25  
 Percent Similarity: 51.22% Conservative: 17  
 Best Local Similarity: 30.49% Mismatches: 31  
 Query Match: 13.85% Indels: 9  
 DB: 1 Gaps: 3

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-07-748-761-2 (1-652)

Qy 22 VallilePheLeuAspValLysHisPheSerProGluAspLeuThrValLysValGlnAsp 41  
 Db 219 GTGTTCAAGTGTGATCTTCAGGGGCTTAAGAGGAGGAAGTCAAAAGTGAAGTCGAGGAG 278  
 Qy 42 AspPheVal---GluileHisGlyLysHisGlnGluArgGlnAspHisGly----- 58  
 Db 279 GATAGGGTTCATGATCAGCGAGAGGAGGAACTGGGAGAGGAAGTAAGTAAGTAAG 338  
 Qy 59 -----TyrIleSerArgGluPheHisArgArgTyrArgLeuProSerAsn 73  
 Db 339 TGGCATCGATCGAGCGAGGAGCGGAAATTCATGAGGAGATTTAGACTTCCGAGAA 398  
 Qy 74 ValaAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGly 93  
 Db 399 GCAAAGATGGATCAAGTTAAGCGCTATGAGG---AATGAGTGCTTACTGTTACTGTT 455  
 Qy 94 ProLys 95  
 Db 456 CCAAAG 461

## RESULT 37

US-09-949-016-13807/c  
 ; Sequence 13807, Application US/09949016  
 ; Patent No. 6812339  
 ; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 13807

; LENGTH: 168971

; TYPE: DNA

; ORGANISM: Human

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)...(168971)

; OTHER INFORMATION: n = A,T,C or G

US-09-949-016-13807

## Alignment Scores:

Pred. No.: 42.1 Length: 168971  
 Score: 88.00 Matches: 30  
 Percent Similarity: 46.99% Conservative: 9  
 Best Local Similarity: 36.14% Mismatches: 30  
 Query Match: 13.77% Indels: 14  
 DB: 4 Gaps: 4

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-09-949-016-13807 (1-168971)

Qy 49 LysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArgArg--- 67  
 Db 167465 AAACCTACTGAGGTGAGGAGGACCGGGA---GCCGGGAGGAGGCGCATCGCAGGAGA 167409

Qy 68 ---TyrArgLeuProSerAsnValAspGlnSerAlaLeu-----SerCys 81  
 Db 167408 AAGTGGCGGCTCCCTCGCTCATTTCCAGGAGCAGCTGGGCGAGTGGTTCGCTCAACTGT 167349

Qy 82 SerLeuSer-----AlaAspGlyMetLeuThrPheCysGlyProLys 95  
 Db 167348 GCTCTGAGCTGGGCTCCAGTGACGGCACTGAGGGGATGGGAGGTCTCTGTGGGGGACTC 167289

Qy 96 IleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProValSerArgGluGlu 115  
 Db 167288 CTACCACCGGCTACCAGCGTGTGTGCCGAGTCCCTCTCTCTCTCTCTCTCTCTCT 167229

Qy 116 LysProThr 118  
 Db 167228 AAGCCCAACC 167220

RESULT 38  
 US-09-902-540-6436  
 ; Sequence 6436, Application US/09902540  
 ; Patent No. 6833447  
 ; GENERAL INFORMATION:

; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Wisegand, Roger C.  
 ; TITLE OF INVENTION: MYCOCCOCUS xanthus Genome Sequences and Uses Thereof  
 ; FILE REFERENCE: 38-10(15849)B  
 ; CURRENT APPLICATION NUMBER: US/09/902,540  
 ; CURRENT FILING DATE: 2001-07-10  
 ; PRIOR APPLICATION NUMBER: 60/217,883  
 ; PRIOR FILING DATE: 2000-07-10  
 ; NUMBER OF SEQ ID NOS: 16825  
 ; SEQ ID NO 6436  
 ; LENGTH: 489  
 ; TYPE: DNA  
 ; ORGANISM: Myxococcus xanthus  
 US-09-902-540-6436

Alignment Scores:  
 Pred. No.: 0.0111 Length: 489  
 Score: 86.00 Matches: 33  
 Percent Similarity: 43.36% Conservative: 16  
 Best Local Similarity: 29.20% Mismatches: 39  
 Query Match: 13.46% Indels: 26  
 DB: 4 Gaps: 4

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-09-902-540-6436 (1-489)

Qy 13 GluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPheSerPro 32  
 Db 154 GAGGTGAAGGAGACCCAGGATGCCTTCATCTTCAAGCCGATGTCCTCCCGCGGTGGAGGAG 213

Qy 33 GluAspLeuThrValLysValGlnAspPheValGluIleHisGlyLysHisAsnGlu 52  
 Db 214 AAGGACCTGGAAATCACCTCGCGAGAACCGGTGACCATCAGCGGCAACGCGGAAGA 273

Qy 53 ArgGlnAspAspHisGly-----TyrIleSerArgGlu-----PheHis 65  
 Db 274 GAGCGCGCGACGAAGGCGACCGCTACTACGCTACGAGCGCAACTATGCTCGTCTCAGC 333

```

Qy 66 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 85
Db 334 CGACCTTCAGCTCCCGGGGTGTG-----AATGCC 366

Qy 86 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 105
Db 367 GAC-----AAGCTCCAGGC-GGACTTCAAGAGTGGCGTCCT 401

Qy 106 GluArgAla-IleProValSerArgGluGluLysPro 117
Db 402 CAACGTGCGCATCCCCAAGAAAGTCCGAGGAGCAGCCC 438

RESULT 39
US-09-902-540-444
; Sequence 444, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 444
; LENGTH: 3626
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(3626)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-444

Alignment Scores:
Pred. No.: 0.234 Length: 3626
Score: 86.00 Matches: 33
Percent Similarity: 43.36% Conservative: 16
Best Local Similarity: 29.20% Mismatches: 39
Query Match: 13.46% Indels: 26
DB: 4 Gaps: 4

US-10-657-740-1_COPY_51_173 (1-123) x US-09-902-540-444 (1-3626)
Qy 13 GluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPheSerPro 32
Db 1509 GAGGTGAGGAGACCCAGGATGCCCTTCATCTTCAAGCGCGGATGTCCTCCCGCGGTGGAGGAG 1568

Qy 33 GluAspLeuThrValLysValGlnAspPheValGluIleHisGlyLysHisAsnGlu 52
Db 1569 AAGGACCTTGAATCACCTCCGCGAAGACCGGCTGACCATCAGCGGCAACGGGAAGAA 1628

Qy 53 ArgGlnAspAspHisGly-----TyrIleSerArgGlu-----PheHis 65
Db 1629 GAGCGCCCGCAGCAAGAGGACCGCTACTACGCTACGAGCGCAACTATGGCTCGTTCAGC 1688

Qy 66 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 85
Db 1689 CGCACTTCACGTCCTCCCGGGGTGTG-----AATGCC 1721

Qy 86 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 105
Db 1722 GAC-----AAGCTCCAGGC-GGACTTCAAGAGTGGCGTCCT 1756

Qy 106 GluArgAla-IleProValSerArgGluGluLysPro 117
Db 1757 CAACGTGCGCATCCCCAAGAAAGTCCGAGGAGCAGCCC 1793

RESULT 40

```

```

US-09-489-039A-3463
; Sequence 3463, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 3463
; LENGTH: 576
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-3463

Alignment Scores:
Pred. No.: 0.0198 Length: 576
Score: 85.00 Matches: 23
Percent Similarity: 40.00% Conservative: 19
Best Local Similarity: 21.90% Mismatches: 43
Query Match: 13.30% Indels: 20
DB: 4 Gaps: 2

US-10-657-740-1_COPY_51_173 (1-123) x US-09-489-039A-3463 (1-576)
Qy 17 AspArgAspLysPheValIlePheLeuAspValLysHisPheSerProGluAspLeuThr 36
Db 289 GACGAAACCACTATCGCATCGCTATCGCGTGGCTTGTCTGAAACGAGCTGGAG 348

Qy 37 ValLysValGlnAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAsp 56
Db 349 ATCACTGCCAGGACCAATCTGCTGCTCAAAGGCCGCCACGCCCGGAGGAGAGAG 408

Qy 57 HisGlyTyrIle-----SerArgGluPheHisArgTyrArgLeuPro 71
Db 409 CGGACCTTACCTGTATCAGGGGATCGCGGAGCGCAACTTCGAGCGCAATTCAGCTGGCG 468

Qy 72 SerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPhe 91
Db 469 GAAACACATTCACGTCGCG-----486

Qy 92 CysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProVal 111
Db 487 ---GGCGCCAATCTGTGTGAACGCGCTGTATATCGATCTGGAACGGGTGATCCCGGAA 543

Qy 112 SerArgGluGluLys 116
Db 544 GCGAACACGCCGCGC 558

RESULT 41
US-09-583-110-2262
; Sequence 2262, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 2262
; LENGTH: 1350

```

```
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-2262

Alignment Scores:
Pred. No.: 0.437 Length: 1350
Score: 79.50 Matches: 25
Percent Similarity: 46.99% Conservative: 14
Best Local Similarity: 30.12% Mismatches: 27
Query Match: 12.44% Indels: 17
DB: 4 Gaps: 4

US-10-657-740-1_COPY_51_173 (1-123) x US-09-583-110-2262 (1-1350)
QY 26 AspVallysHisPheSerProGluAspLeuThrVallysValGln---AspAspPheVal 44
DB 67 GACATCAAGGCCAACAGCTCAAGTAATGTGTGATGCTCTTCAAGAGCTGATCAGTTG 126
QY 45 GluIleHisGlyLyHisAsnGluArgGlnAspAspHisGlyTyr----- 59
DB 127 AGTATTACCTTAAGCAACAGATTTTCCCAAGAGCATGAACCTCTTAAAGATATCAAG 186
QY 60 IleSerArgGluPheHisArgTyrArgLeuProSerAsnValAsp----- 75
DB 187 ATCGGCAACACTCTTTTAAAGAGGTGAGCTCCCTTCTAAGTTGATTCAGTCGTAAAA 246
QY 76 -----GlnSerAlaLeuSerCysSerLeuSerAla-----AspGlyMet 88
DB 247 GTTTACTTTGAAAGTGTGTAGTGTCTTCTCTAGCCAGGCGATGCTGCGCGCATG 306
QY 89 LeuThrPhe 91
DB 307 GAAACTTTT 315

RESULT 42
US-09-107-433-2322
; Sequence 2322, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGN
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 2322:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1353 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...1353
; SEQUENCE DESCRIPTION: SEQ ID NO: 2322:
US-09-107-433-2322

Alignment Scores:
Pred. No.: 0.438 Length: 1353
Score: 79.50 Matches: 25
Percent Similarity: 46.99% Conservative: 14
Best Local Similarity: 30.12% Mismatches: 27
Query Match: 12.44% Indels: 17
DB: 4 Gaps: 4

US-10-657-740-1_COPY_51_173 (1-123) x US-09-107-433-2322 (1-1353)
QY 26 AspVallysHisPheSerProGluAspLeuThrVallysValGln---AspAspPheVal 44
DB 70 GACATCAAGGCCAACAGCTCAAGTAATGTGTGATGCTCTTCAAGAGCTGATCAGTTG 129
QY 45 GluIleHisGlyLyHisAsnGluArgGlnAspAspHisGlyTyr----- 59
DB 130 AGTATTACCTTAAGCAACAGATTTTCCCAAGAGCATGAACCTCTTAAAGATATCAAG 189
QY 60 IleSerArgGluPheHisArgTyrArgLeuProSerAsnValAsp----- 75
DB 190 ATCGGCAACACTCTTTTAAAGAGGTGAGCTCCCTTCTAAGTTGATTCAGTCGTAAAA 249
QY 76 -----GlnSerAlaLeuSerCysSerLeuSerAla-----AspGlyMet 88
DB 250 GTTTACTTTGAAAGTGTGTAGTGTCTTCTCTAGCCAGGCGATGCTGCGCGCATG 309
QY 89 LeuThrPhe 91
DB 310 GAAACTTTT 318

RESULT 43
US-09-949-016-11770/C
; Sequence 11770, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11770
; LENGTH: 49487
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(49487)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-11770
```

## Alignment Scores:

Pred. No.: 202 Length: 49487  
Score: 77.50 Matches: 41  
Percent Similarity: 39.53% Conservative: 27  
Best Local Similarity: 23.84% Mismatches: 53  
Query Match: 12.13% Indels: 51  
DB: 4 Gaps: 10

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-09-949-016-11770 (1-49487)

Qy 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAsp----- 17  
Db 2663 TCTCTGATTCTAGGAGTTTCTTCACCGTTTAAGTGAGCTGAGGTGTAACGACCAT 2604  
Qy 18 -----ArgAspLysPheValIlePheLeuAsp----- 26  
Db 2603 AGATAATATCCAGATCGGAAACCCCGAGAGCGCAGGATTATCTGGATAACGGAGCG 2544  
Qy 27 -----ValLysHisPheSerProGlu-----AspLeuThr 36  
Db 2543 GATGGAGGATTGCTCCGTTCTCCACACTCAAGCCCGAAAGGTCACTGACCTAGTC 2484  
Qy 37 ValLysValGlnAspPheValGluIleHisGlyLys---HisAsn----- 51  
Db 2483 TTATTCTTTACAGAGAAATCGGCAACCCATGCGCGCTCCACAATCTCTATAGCGCG 2424  
Qy 52 -----GluArgGlnAsp-----AspHisGlyTyr 59  
Db 2423 CGACTCGGTGTACAGGGAAGCGGGAACACACACACTTCTACGCAGACCGGTTAC 2364  
Qy 60 IleSerArgGluPheHisArgTyrArgLeuProSerAsnVal----- 74  
Db 2363 ATTAAGCGCGACGCCACACACACCGCCACACACACACACACACACACACACAC 2304  
Qy 75 AspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyPro 94  
Db 2303 CCCACACTTCGTTAACTTGTGATGATAAATTAGGCGCAGGTGACACTACTGT---CCC 2247  
Qy 95 LysIleGlnThr-----GlyLeuAspAlaThrHisAlaGluArgAlaIleProVal 111  
Db 2246 GAACCTCAGCTGGAACGCTAGTCCAGTCTCTCACTGCTCACTGACGTAGCAGTCCGACC 2187  
Qy 112 SerArgGluGluLysProThrSerAlaProSerSer 123  
Db 2186 TTGTCCGAGGAGTTTCT---TCCGCTCCCTCACT 2154

## RESULT 44

US-09-949-016-13631  
; Sequence 13631, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13631  
; LENGTH: 57751  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-13631

## Alignment Scores:

Pred. No.: 0.97 Length: 966  
Score: 75.50 Matches: 28  
Percent Similarity: 41.84% Conservative: 13  
Best Local Similarity: 28.57% Mismatches: 40  
Query Match: 11.82% Indels: 18  
DB: 4 Gaps: 5

Pred. No.: 255 Length: 57751  
Score: 77.50 Matches: 41  
Percent Similarity: 39.53% Conservative: 27  
Best Local Similarity: 23.84% Mismatches: 53  
Query Match: 12.13% Indels: 51  
DB: 4 Gaps: 10

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-09-949-016-13631 (1-57751)

Qy 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAsp----- 17  
Db 1009 TCTCTGATTCTAGGAGTTTCTTCACCGTTTAAGTGAGCTGAGGTGTAACGACCAT 1068  
Qy 18 -----ArgAspLysPheValIlePheLeuAsp----- 26  
Db 1069 AGATAATATCCAGATCGGAAACCCCGAGAGCGCAGGATTATCTGGATAACGGAGCG 1128  
Qy 27 -----ValLysHisPheSerProGlu-----AspLeuThr 36  
Db 1129 GATGGAGGATTGCTCCGTTCTCCACACTCAAGCCCGAAAGGTCACTGACCTAGTC 1188  
Qy 37 ValLysValGlnAspPheValGluIleHisGlyLys---HisAsn----- 51  
Db 1189 TTATTCTTTACAGAGAAATCGGCAACCCATGCGCGCTCCACAATCTCTATAGCGCG 1248  
Qy 52 -----GluArgGlnAsp-----AspHisGlyTyr 59  
Db 1249 CGACTCGGTGTACAGGGAAGCGGGAACACACACACTTCTACGCAGACCGGTTAC 1308  
Qy 60 IleSerArgGluPheHisArgTyrArgLeuProSerAsnVal----- 74  
Db 1309 ATTAAGCGCGACGCCACACACCGCCACACACACACACACACACACACACACAC 1368  
Qy 75 AspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyPro 94  
Db 1369 CCCACACTTCGTTAACTTGTGATGATAAATTAGGCGCAGGTGACACTACTGT---CCC 1425  
Qy 95 LysIleGlnThr-----GlyLeuAspAlaThrHisAlaGluArgAlaIleProVal 111  
Db 1426 GAACCTCAGCTGGAACGCTAGTCCAGTCTCTCACTGCTCACTGACGTAGCAGTCCGACC 1485  
Qy 112 SerArgGluGluLysProThrSerAlaProSerSer 123  
Db 1486 TTGTCCGAGGAGTTTCT---TCCGCTCCCTCACT 1518

## RESULT 45

US-09-969-532-23  
; Sequence 23, Application US/09969532  
; Patent No. 6777232  
; GENERAL INFORMATION:

; APPLICANT: Walke, D. Wade  
; APPLICANT: Scoville, John  
; TITLE OF INVENTION: No. 6777232el Human Membrane Proteins and Polynucleotides Encodin  
; FILE REFERENCE: LEX-0244-USA  
; CURRENT APPLICATION NUMBER: US/09/969,532  
; CURRENT FILING DATE: 2001-10-02  
; PRIOR APPLICATION NUMBER: US 60/237,280  
; PRIOR FILING DATE: 2000-10-02  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 23  
; LENGTH: 966  
; TYPE: DNA  
; ORGANISM: homo sapiens  
US-09-969-532-23

## Alignment Scores:

Pred. No.: 0.97 Length: 966  
Score: 75.50 Matches: 28  
Percent Similarity: 41.84% Conservative: 13  
Best Local Similarity: 28.57% Mismatches: 40  
Query Match: 11.82% Indels: 18  
DB: 4 Gaps: 5

```
US-10-657-740-1_COPY_51_173 (1-123) x US-09-969-532-23 (1-966)
QY 30 PheSerPro---GluAspLeuThrVallysValGlnAspAspPheVal----- 44
Db 529 TTTAAACCCCTTTCGACATCAAGAGTGAAGTCCAGAGCTCGTTTCATGTTCCCTCGGA 588
QY 45 -----GluIleHisGlyLysHisAsnGluArgGlnAspHisGly--- 58
Db 589 GTGCTGAGAGAGCTGAGTACCACGCGCAAGAATCATTCAGGACTTTTCCCATGGAAC 648
QY 59 ---TyrIleSerArgGluPheHisArgArgTyrArgLeuPro-----SerAsnValAsp 75
Db 649 ACCACAGCTTTAGTACATGATCCAGAAATATAATGCCCTACATCCAAATCTGTCA 708
QY 76 GlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLys 95
Db 709 TCACCTCCCAAGGACAGCAACTGAGGACAACTGGTGTCTTTGGCCATTTAGGGGGCGC 768
QY 96 Ile-----GlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 110
Db 769 TTAGTAATGCCAAATACAGGGGT-GAGCTTACTCATACACACGCGTGCCATCCC 821

RESULT 46
US-09-969-532-21
; Sequence 21, Application US/09969532
; Patent No. 6777232
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; TITLE OF INVENTION: No. 6777232el Human Membrane Proteins and Polynucleotides Encodir
; FILE REFERENCE: LEX-0244-USA
; CURRENT APPLICATION NUMBER: US/09/969,532
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US 60/237,280
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq For Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 999
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-969-532-21

Alignment Scores:
Pred. No.: 1,02 Length: 999
Score: 75.50 Matches: 28
Percent Similarity: 41.84% Conservatve: 13
Best Local Similarity: 28.57% Mismatches: 40
Query Match: 11.82% Indels: 18
DB: 4 Gaps: 5

US-10-657-740-1_COPY_51_173 (1-123) x US-09-969-532-21 (1-999)
QY 30 PheSerPro---GluAspLeuThrVallysValGlnAspAspPheVal----- 44
Db 562 TTTAAACCCCTTTCGACATCAAGAGTGAAGTCCAGAGCTCGTTTCATGTTCCCTCGGA 621
QY 45 -----GluIleHisGlyLysHisAsnGluArgGlnAspHisGly--- 58
Db 622 GTGCTGAGAGAGCTGAGTACCACGCGCAAGAATCATTCAGGACTTTTCCCATGGAAC 681
QY 59 ---TyrIleSerArgGluPheHisArgArgTyrArgLeuPro-----SerAsnValAsp 75
Db 682 ACCACAGCTTTAGTACATGATCCAGAAATATAATGCCCTACATCCAAATCTGTCA 741
QY 76 GlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLys 95
Db 742 TCACCTCCCAAGGACAGCAACTGAGGACAACTGGTGTCTTTGGCCATTTAGGGGGCGC 801
QY 96 Ile-----GlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 110
Db 802 TTAGTAATGCCAAATACAGGGGT-GAGCTTACTCATACACACGCGTGCCATCCC 854
```

```
RESULT 47
US-09-969-532-19
; Sequence 19, Application US/09969532
; Patent No. 6777232
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; TITLE OF INVENTION: No. 6777232el Human Membrane Proteins and Polynucleotides Encodir
; FILE REFERENCE: LEX-0244-USA
; CURRENT APPLICATION NUMBER: US/09/969,532
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US 60/237,280
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq For Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 1008
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-969-532-19

Alignment Scores:
Pred. No.: 1,04 Length: 1008
Score: 75.50 Matches: 28
Percent Similarity: 41.84% Conservatve: 13
Best Local Similarity: 28.57% Mismatches: 40
Query Match: 11.82% Indels: 18
DB: 4 Gaps: 5

US-10-657-740-1_COPY_51_173 (1-123) x US-09-969-532-19 (1-1008)
QY 30 PheSerPro---GluAspLeuThrVallysValGlnAspAspPheVal----- 44
Db 571 TTTAAACCCCTTTCGACATCAAGAGTGAAGTCCAGAGCTCGTTTCATGTTCCCTCGGA 630
QY 45 -----GluIleHisGlyLysHisAsnGluArgGlnAspHisGly--- 58
Db 631 GTGCTGAGAGAGCTGAGTACCACGCGCAAGAATCATTCAGGACTTTTCCCATGGAAC 690
QY 59 ---TyrIleSerArgGluPheHisArgArgTyrArgLeuPro-----SerAsnValAsp 75
Db 691 AACACAGCTTTAGTACATGATCCAGAAATATAATGCCCTACATCCAAATCTGTCA 750
QY 76 GlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLys 95
Db 751 TCACCTCCCAAGGACAGCAACTGAGGACAACTGGTGTCTTTGGCCATTTAGGGGGCGC 810
QY 96 Ile-----GlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 110
Db 811 TTAGTAATGCCAAATACAGGGGT-GAGCTTACTCATACACACGCGTGCCATCCC 863

RESULT 48
US-09-969-532-17
; Sequence 17, Application US/09969532
; Patent No. 6777232
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; TITLE OF INVENTION: No. 6777232el Human Membrane Proteins and Polynucleotides Encodir
; FILE REFERENCE: LEX-0244-USA
; CURRENT APPLICATION NUMBER: US/09/969,532
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US 60/237,280
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq For Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 1041
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-969-532-17
```

## Alignment Scores:

Pred. No.: 1.09 Length: 1041  
Score: 75.50 Matches: 28  
Percent Similarity: 41.84% Conservative: 13  
Best Local Similarity: 28.57% Mismatches: 40  
Query Match: 11.82% Indels: 18  
DB: 4 Gaps: 5

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-09-969-532-17 (1-1041)

QY 30 PheSerPro---GluAspLeuThrValGlnAspPheVal----- 44  
Db 604 TTTAACCCCTTTGTCGGACATCAAGTGAAGTCCAGAGCTCGTTTCATGTTCCCTGGGA 663  
QY 45 -----GluIleHisGlyLysHisAsnGluArgGlnAspHisGly--- 58  
Db 664 GTGCTGAGAGAGCTGAGTACACGGCAGAGATCATTTCCAGGACTTTTCCCATGGAAAC 723  
QY 59 ---TyrIleSerArgGluPheHisArgTyrArgLeuPro-----SerAsnValAsp 75  
Db 724 AACACAGCTTTAGTACAAATGCATCCAGAAATAAATGACCTACATCCAAATCTGTCA 783  
QY 76 GlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLys 95  
Db 784 TCACTCCCCACAGCAGACAGACTGAGGACAACTGGTGTCTTTGGCCATTTAGGGGGCGC 843  
QY 96 Ile-----GlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 110  
Db 844 TTAGTAATGCCAAATACAGGGGT-GAGCTTACTATACACACGGTGCCATCCC 896

## RESULT 49

US-09-969-532-7  
; Sequence 7, Application US/09969532  
; Patent No. 6777232  
; GENERAL INFORMATION:  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Scoville, John  
; TITLE OF INVENTION: No. 6777232el Human Membrane Proteins and Polynucleotides Encodin  
; FILE REFERENCE: LEX-0244-USA  
; CURRENT APPLICATION NUMBER: US/09/969,532  
; PRIOR FILING DATE: 2001-10-02  
; PRIOR APPLICATION NUMBER: US 60/237,280  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7  
; TYPE: DNA  
; ORGANISM: homo sapiens  
US-09-969-532-7

## Alignment Scores:

Pred. No.: 2.21 Length: 1659  
Score: 75.50 Matches: 28  
Percent Similarity: 41.84% Conservative: 13  
Best Local Similarity: 28.57% Mismatches: 40  
Query Match: 11.82% Indels: 18  
DB: 4 Gaps: 5

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-09-969-532-7 (1-1659)

QY 30 PheSerPro---GluAspLeuThrValGlnAspPheVal----- 44  
Db 1222 TTTAACCCCTTTGTCGGACATCAAGTGAAGTCCAGAGCTCGTTTCATGTTCCCTGGGA 1281  
QY 45 -----GluIleHisGlyLysHisAsnGluArgGlnAspHisGly--- 58  
Db 1282 GTGCTGAGAGAGCTGAGTACACGGCAGAGATCATTTCCAGGACTTTTCCCATGGAAAC 1341  
QY 59 ---TyrIleSerArgGluPheHisArgTyrArgLeuPro-----SerAsnValAsp 75  
Db 1342 AACACAGCTTTAGTACAAATGCCAGAAATAAATGACCTACATCCAAATCTGTCA 1401

QY 76 GlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLys 95  
Db 1402 TCACTCCCCACAGCAGACACTGGTGCTTTTGGCCATTTAGGGGGCGC 1461  
QY 96 Ile-----GlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 110  
Db 1462 TTAGTAATGCCAAATACAGGGGT-GAGCTTACTATACACACGGTGCCATCCC 1514

## RESULT 50

US-09-969-532-5  
; Sequence 5, Application US/09969532  
; Patent No. 6777232  
; GENERAL INFORMATION:  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Scoville, John  
; TITLE OF INVENTION: No. 6777232el Human Membrane Proteins and Polynucleotides Encodin  
; FILE REFERENCE: LEX-0244-USA  
; CURRENT APPLICATION NUMBER: US/09/969,532  
; PRIOR FILING DATE: 2001-10-02  
; PRIOR APPLICATION NUMBER: US 60/237,280  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 1692  
; TYPE: DNA  
; ORGANISM: homo sapiens  
US-09-969-532-5

## Alignment Scores:

Pred. No.: 2.28 Length: 1692  
Score: 75.50 Matches: 28  
Percent Similarity: 41.84% Conservative: 13  
Best Local Similarity: 28.57% Mismatches: 40  
Query Match: 11.82% Indels: 18  
DB: 4 Gaps: 5

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-09-969-532-5 (1-1692)

QY 30 PheSerPro---GluAspLeuThrValGlnAspPheVal----- 44  
Db 1255 TTTAACCCCTTTGTCGGACATCAAGTGAAGTCCAGAGCTCGTTTCATGTTCCCTGGGA 1314  
QY 45 -----GluIleHisGlyLysHisAsnGluArgGlnAspHisGly--- 58  
Db 1315 GTGCTGAGAGAGCTGAGTACACGGCAGAAATCATTCAGGACTTTTCCCATGGAAAC 1374  
QY 59 ---TyrIleSerArgGluPheHisArgTyrArgLeuPro-----SerAsnValAsp 75  
Db 1375 AACACAGCTTTAGTACAAATGACATCCAGAAATAAATGACCTACATCCAAATCTGTCA 1434  
QY 76 GlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLys 95  
Db 1435 TCACTCCCCACAGCAGACACTGAGGACAACTGGTGTCTTTGGCCATTTAGGGGGCGC 1494  
QY 96 Ile-----GlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 110  
Db 1495 TTAGTAATGCCAAATACAGGGGT-GAGCTTACTATACACACGGTGCCATCCC 1547

Search completed: May 30, 2005, 08:54:27  
Job time : 171.922 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 30, 2005, 05:14:41 ; Search time 864.74 Seconds  
(without alignments)  
873.287 Million cell updates/sec

Title: US-10-657-740-1\_COPY\_51\_173

Perfect score: 639

Sequence: 1 SLFRVLDSGISEVRSRDK.....HAERAIPVSRREKPTAPSS 123

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 5700845 seqs, 3069779757 residues

Total number of hits satisfying chosen parameters: 11401690

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO\_spool/US10657740/runat\_27052005\_165254\_3380/app\_query.fasta\_1.590  
-DB=Published Applications NA -OPMT=fastap -SUFFIX=rnnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITIS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=150 -LOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=50 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US10657740 @CGN 1 904 @runat\_27052005\_165254\_3380  
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2\_6/ptodata/1/pubna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/1/pubna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/1/pubna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/1/pubna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/1/pubna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/1/pubna/US10C\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/1/pubna/US10D\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/1/pubna/US10E\_PUBCOMB.seq.\*  
18: /cgn2\_6/ptodata/1/pubna/US10F\_PUBCOMB.seq.\*  
19: /cgn2\_6/ptodata/1/pubna/US10\_NEW\_PUB.seq.\*  
20: /cgn2\_6/ptodata/1/pubna/US11\_NEW\_PUB.seq.\*  
21: /cgn2\_6/ptodata/1/pubna/US60\_NEW\_PUB.seq.\*  
22: /cgn2\_6/ptodata/1/pubna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	639	100.0	372	18	US-10-657-740-2	Sequence 2, Appli
2	599	93.7	1056	16	US-10-316-253-112	Sequence 112, App
3	577.5	90.4	1271	16	US-10-316-253-110	Sequence 110, App
4	567	88.7	531	13	US-10-105-427-1	Sequence 1, Appli
5	367	57.4	211	16	US-10-029-386-26581	Sequence 26581, A
6	367	57.4	573	16	US-10-029-386-12881	Sequence 12881, A
7	361	56.5	528	17	US-10-152-319A-1574	Sequence 1574, Ap
8	361	56.5	689	19	US-10-486-706-455	Sequence 455, App
9	361	56.5	1247	9	US-09-917-800A-1419	Sequence 1419, Ap
10	361	56.5	1247	19	US-10-486-706-105	Sequence 105, App
11	358	56.0	548	15	US-10-101-510-250	Sequence 250, App
12	354	55.4	528	18	US-10-770-668-35	Sequence 35, Appli
13	354	55.4	691	9	US-09-954-456-514	Sequence 514, App
14	354	55.4	691	10	US-09-960-706-869	Sequence 869, App
15	354	55.4	691	10	US-09-873-319-566	Sequence 566, App
16	354	55.4	691	10	US-09-873-367C-764	Sequence 764, App
17	354	55.4	691	17	US-10-133-937-61	Sequence 61, Appli
18	354	55.4	691	17	US-10-172-118-686	Sequence 686, App
19	354	55.4	691	17	US-10-159-563-61	Sequence 61, Appli
20	354	55.4	691	17	US-10-342-887-686	Sequence 686, App
21	354	55.4	691	19	US-10-843-641A-764	Sequence 764, App
22	354	55.4	691	19	US-10-843-641A-3541	Sequence 3541, App
23	354	55.4	691	19	US-10-367-057-187	Sequence 187, App
24	332.5	52.0	380	9	US-09-960-352-4143	Sequence 4143, Ap
25	283	44.3	449	9	US-09-960-352-4277	Sequence 4277, Ap
26	282.5	44.2	373	17	US-10-242-535A-26398	Sequence 26398, A
27	282.5	44.2	373	17	US-10-085-783A-26398	Sequence 26398, A
28	263	41.2	349	9	US-09-990-747-28	Sequence 28, Appli
29	261.5	40.9	764	18	US-10-605-498-91	Sequence 91, Appli
30	261.5	40.9	789	17	US-10-641-643-1300	Sequence 1300, App
31	261.5	40.9	847	14	US-10-153-668-283	Sequence 283, App
32	261.5	40.9	865	17	US-09-969-034-4480	Sequence 4480, Ap
33	261.5	40.9	865	17	US-10-172-118-626	Sequence 626, App
34	261.5	40.9	865	17	US-10-342-887-626	Sequence 626, App
35	261.5	40.9	1231	9	US-09-880-107-3865	Sequence 3865, Ap
36	261.5	40.9	1380	14	US-10-100-957A-169	Sequence 169, App
37	261	40.8	604	9	US-09-917-800A-981	Sequence 981, App
38	260	40.7	599	14	US-10-060-036-43	Sequence 43, Appli
39	260	40.7	634	19	US-10-784-420-43	Sequence 43, Appli
40	260	40.7	787	9	US-09-917-800A-1428	Sequence 1428, Ap
41	260	40.7	787	17	US-10-191-803-228	Sequence 228, App
42	260	40.7	787	17	US-10-152-319A-1963	Sequence 1963, Ap
43	259	40.5	396	9	US-09-960-352-12622	Sequence 12622, A
44	250.5	39.2	393	18	US-10-425-115-63403	Sequence 63403, A
45	250.5	39.2	480	14	US-10-226-956-320	Sequence 320, App
46	250.5	39.2	486	14	US-10-226-956-303	Sequence 303, App
47	250.5	39.2	1310	17	US-10-152-319A-1412	Sequence 1412, App
48	250.5	39.2	1820	18	US-10-723-860-5075	Sequence 5075, Ap
49	250	39.1	450	10	US-09-918-995-4949	Sequence 4949, Ap
50	248	38.8	503	10	US-09-911-504-146	Sequence 146, App
51	242	37.9	543	17	US-10-621-901-2279	Sequence 2279, Ap
52	241.5	37.8	708	10	US-09-814-353-14770	Sequence 14770, A
53	240.5	37.6	461	14	US-10-198-846-10360	Sequence 10360, A
54	240	37.6	563	14	US-10-198-846-8376	Sequence 8376, Ap
55	234.5	36.7	614	10	US-09-814-353-2040	Sequence 2040, Ap
56	234.5	36.7	614	10	US-09-814-353-8386	Sequence 8386, Ap
57	233.5	36.5	954	10	US-09-814-353-20273	Sequence 20273, A
58	229	35.8	1700	9	US-09-813-358-2	Sequence 2, Appli
59	229	35.8	1700	10	US-09-997-279-2	Sequence 2, Appli
60	226	35.4	421	10	US-09-918-995-4772	Sequence 4772, Ap
61	223.5	35.0	500	10	US-09-918-995-22134	Sequence 22134, A
62	222	34.7	405	10	US-09-918-995-4075	Sequence 4075, Ap
63	217	34.0	605	19	US-10-840-038-2	Sequence 2, Appli
64	217	34.0	736	19	US-10-840-038-3	Sequence 3, Appli
65	212	33.2	49634	11	US-09-968-007A-514	Sequence 514, App
66	212	33.2	49634	19	US-10-843-641A-6984	Sequence 6984, App
67	210.5	32.9	829	17	US-10-062-674-1924	Sequence 1924, Ap
68	209.5	32.8	874	17	US-10-133-937-62	Sequence 62, Appli
69	209.5	32.8	874	17	US-10-159-563-62	Sequence 62, Appli

70	32.4	387	18	US-10-723-860-380	Sequence 380, App
71	202	212	18	US-10-482-158-43	Sequence 43, Appl
72	199	381	18	US-10-425-115-173881	Sequence 173881, App
c 73	198.5	396	9	US-09-954-531-845	Sequence 845, App
c 74	198.5	396	9	US-09-954-531-1265	Sequence 1265, App
c 75	198.5	396	19	US-10-843-641A-1912	Sequence 1912, App
c 76	198.5	396	19	US-10-843-641A-2332	Sequence 2332, App
c 77	197.5	396	19	US-10-843-641A-2332	Sequence 2332, App
78	195	349	10	US-09-991-936-341	Sequence 341, App
79	192.5	467	14	US-10-060-036-34	Sequence 34, Appl
80	192.5	482	10	US-09-918-995-4433	Sequence 4433, App
81	192	550	17	US-10-621-901-2194	Sequence 2194, App
82	188	4206	15	US-10-205-331-105	Sequence 105, App
83	188	4206	15	US-10-101-510-288	Sequence 288, App
84	182.5	4206	18	US-10-283-975A-93	Sequence 93, Appl
c 85	179	565	14	US-10-468-091-35	Sequence 35, Appl
c 86	174	28.0	510	US-10-198-846-366	Sequence 366, App
c 87	174	27.2	510	US-10-363-345A-653	Sequence 653, App
c 88	174	27.2	510	US-10-363-345A-654	Sequence 654, App
c 89	174	27.2	510	US-10-363-483A-654	Sequence 654, App
90	172	542	17	US-10-363-483A-654	Sequence 654, App
91	170	26.9	542	US-10-621-901-2146	Sequence 2146, App
92	165.5	654	19	US-10-204-921-16	Sequence 16, Appl
93	165.5	487	17	US-10-369-493-30110	Sequence 30110, A
94	163	487	17	US-10-369-493-30111	Sequence 30111, A
95	160	445	10	US-09-918-995-10509	Sequence 10509, A
96	159.5	1495	16	US-10-006-285-269	Sequence 269, App
97	159	24.9	1627	US-10-062-674-254	Sequence 254, App
98	159	24.9	2010	US-10-116-802-322	Sequence 322, App
99	159	24.9	2010	US-10-172-118-1482	Sequence 1482, App
100	159	24.9	2036	US-10-342-887-1482	Sequence 447, App
101	158	24.7	810	US-10-006-285-447	Sequence 4925, App
102	148	23.2	403	US-10-723-860-4925	Sequence 3948, App
103	148	23.2	490	US-09-960-352-3948	Sequence 30109, A
c 104	148	23.2	490	US-10-369-493-30109	Sequence 30112, A
c 105	148	23.2	6806	US-10-205-331-105	Sequence 105, App
c 106	142.5	22.3	510	US-10-363-345A-655	Sequence 655, App
c 107	142.5	22.3	510	US-10-363-345A-656	Sequence 656, App
c 108	142.5	22.3	510	US-10-363-483A-655	Sequence 655, App
c 109	142.5	22.3	510	US-10-363-483A-655	Sequence 656, App
c 110	142	22.2	131	US-10-029-386-26899	Sequence 28899, A
c 111	142	22.2	451	US-10-103-524-483	Sequence 483, App
c 112	142	22.2	534	US-10-023-386-13199	Sequence 13199, A
113	137.5	21.5	450	US-09-880-107-3022	Sequence 3022, App
114	137.5	21.5	450	US-09-954-531-962	Sequence 962, App
115	137.5	21.5	450	US-09-954-531-1355	Sequence 1355, App
116	137.5	21.5	450	US-10-843-641A-2029	Sequence 2029, App
117	137.5	21.5	450	US-10-843-641A-2422	Sequence 2422, App
118	135.5	21.2	2632	US-10-044-090-312	Sequence 312, App
119	131	20.5	303	US-09-960-352-2010	Sequence 2010, App
120	131	20.5	395	US-09-918-995-4376	Sequence 4376, App
121	128	20.0	446	US-09-960-352-5348	Sequence 5348, App
122	127.5	20.0	2153	US-10-062-674-2106	Sequence 2106, App
123	125.5	19.6	1891	US-10-316-253-108	Sequence 108, App
124	123	19.2	339	US-09-867-701-6118	Sequence 6118, App
125	121	18.9	65	US-09-908-975-30348	Sequence 30348, A
c 126	120.5	18.9	42339	US-10-087-192-991	Sequence 991, App
127	118	18.5	442	US-09-918-995-5137	Sequence 5137, App
128	118	18.5	495	US-10-255-536-69	Sequence 69, Appl
129	115.5	18.1	478	US-09-918-995-20604	Sequence 20604, A
130	115	18.0	1922	US-10-104-047-1686	Sequence 1686, App
c 131	115	18.0	50002	US-10-087-192-994	Sequence 994, App
132	111.5	17.4	468	US-09-938-842A-2558	Sequence 2558, App
133	111.5	17.4	468	US-09-938-842A-2558	Sequence 2558, App
134	110.5	17.3	719	US-10-037-270-669	Sequence 669, App
135	110.5	17.3	719	US-10-117-722-669	Sequence 669, App
136	110.5	17.3	2152	US-10-723-860-395	Sequence 395, App
137	110.5	17.3	2410	US-10-104-047-972	Sequence 972, App
138	110.5	17.3	3503	US-10-723-860-5087	Sequence 5087, App
c 139	102.5	16.0	583	US-10-021-323-6441	Sequence 6441, App
c 140	102.5	16.0	597	US-10-021-323-5926	Sequence 5926, App
141	102.5	16.0	813	US-10-767-795-5171	Sequence 5171, App
142	101.5	15.9	943	US-10-310-154-324	Sequence 324, App
143	101.5	15.9	943	US-10-732-923-265	Sequence 265, App
144	101.5	15.9	1018	US-10-424-599-103855	Sequence 103855, A
145	101	15.8	60	US-09-908-975-15945	Sequence 15945, A
c 146	101	15.8	2731748	US-10-297-465A-1	Sequence 1, Appli
147	100.5	15.7	473	US-09-918-995-5122	Sequence 5122, App
148	99	15.5	768	US-10-425-114-9506	Sequence 9506, App
149	99	15.5	800	US-10-424-599-18307	Sequence 18307, A
150	98.5	15.4	599	US-10-021-323-5328	Sequence 5328, App
ALIGNMENTS					
RESULT 1					
US-10-657-740-2					
; Sequence 2, Application US/10657740					
; Publication No. US20040157289A1					
GENERAL INFORMATION:					
; APPLICANT: Salerno, John C.					
; APPLICANT: Hanna, Michael					
; APPLICANT: Koretz, Jane F.					
; APPLICANT: Crone, Donna					
; APPLICANT: Smith, Susan E.					
; TITLE OF INVENTION: PROTEIN EXPRESSION SYSTEM					
; FILE REFERENCE: 01794100406US1					
; CURRENT APPLICATION NUMBER: US/10/657,740					
; CURRENT FILING DATE: 2003-09-08					
; PRIOR APPLICATION NUMBER: US 60/408,680					
; PRIOR FILING DATE: 2002-09-06					
; NUMBER OF SEQ ID NOS: 18					
; SOFTWARE: PatentIn version 3.1					
; SEQ ID NO 2					
; LENGTH: 372					
; TYPE: DNA					
; ORGANISM: Homo sapiens					
US-10-657-740-2					
Alignment Scores:					
Pred. No.:	7,71e-84	Length:	372		
Score:	639.00	Matches:	123		
Percent Similarity:	100.00%	Conservative:	0		
Best Local Similarity:	100.00%	Mismatches:	0		
Query Match:	100.00%	Indels:	0		
DB:	18	Gaps:	0		
US-10-657-740-1_COPY_51_173 (1-123) x US-10-657-740-2 (1-372)					
Qy	1	SeqLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys	20		
Db	1	TCCTCTTCGGCACCCTGCTGGACTCCGCATCTCTGAGTTTCGATCCGACCGGACAG	60		
Qy	21	PheValIlePheLeuAspValIleHisPheSerProGluAspLeuThrValLysValGln	40		
Db	61	TTGTCATCTTCTCGATGTGAAGCACTTCTCCCGAGGAGCCTCACCGTGAAGGTGCAG	120		
Qy	41	AspAspPheValIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIle	60		
Db	121	GACGACTTTGTGAGATCCACGGAAGCACACGAGCCGAGGAGCAGCACCGCTACATT	180		
Qy	61	SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer	80		
Db	181	TCCCGTGAATTCACCGCGCTTACCGCTTCCGCTGCGGAGGAGGAGGAGGAGGAGGAG	240		
Qy	81	CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu	100		
Db	241	TGCTCCCTGTCTGCGATGGATGCTGACCTTCTGTGCCCCCAAGATCCAGACTGGCGTG	300		
Qy	101	AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla	120		
Db	301	GATGCCACCCACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	360		
Qy	121	ProSerSer	123		
Db	361	CCCTGCTCC	369		



## RESULT 2

```

US-10-316-253-112
; Sequence 112, Application US/10316253
; Publication No. US20030162706A1
; GENERAL INFORMATION:
; APPLICANT: The Procter & Gamble Company
; APPLICANT: Peters, Kevin
; APPLICANT: Thompson, Larry
; APPLICANT: Wang, Feng
; APPLICANT: Greis, Kenneth
; TITLE OF INVENTION: Angiogenesis Modulating Proteins
; FILE REFERENCE: 8865M
; CURRENT APPLICATION NUMBER: US/10/316,253
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: US 60/355,295
; PRIOR FILING DATE: 2002-02-08
; NUMBER OF SEQ ID NOS: 308
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 112
; LENGTH: 1056
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (13)..(534)
; OTHER INFORMATION:
US-10-316-253-112

```

**Alignment Scores:**

Alignment Scores:	
Pred. No.:	2.67e-77
Score:	599.00
Length:	1056
Matches:	115
Percent Similarity:	79.56%
Conservative:	5
Best Local Similarity:	93.50%
Mismatches:	3
Query Match:	93.74%
Indels:	0
DB:	16
Gaps:	0

US-10-657-740-1 COPY 51 173 (1-123) x US-10-316-253-112 (1-1056)

QY	1	SerLeuPheA	gThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys	20
Db	163	TCCTCTTTCCG	CACAGTGTTCGATCCGGCATCTCTGAGTCCGATCTGACCGGGCAAG	222
QY	21	PheValIlePheLeuAspValIleHisPheSerProGluuAspLeuThrValIysValGln	40	
Db	223	TTTGTCATCTTCTTGGAGTGAAGCACTTCTCTCTGAGACCTCACCGTGAAGGTA	282	
QY	41	AspAspPheValGluIleHisGlyIleHisGlyAsnGluArgGlnAspAspHisGlyIle	60	
Db	283	GAAGATTTTCGTGGAGATCCATGGCCAAACACACAGAGAGGCAGATGACCATGGCTACATT	342	
QY	61	SerArgGluPheHisArgArTyArgLeuProSerAsnValAspGlnSerAlaLeuSer	80	
Db	343	TCCCGTGAATTTACCGTCGCTACCGTCTGCCTTCCATGTGACCAGTCCGCGCTCTCC	402	
QY	81	CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProIysIleGlnThrGlyLeu	100	
Db	403	TGCTCCTTGTCTGCGGATGGCATGTCACCTTCTCTGGCCCCAAGGTGCAGTCTGCCTTG	462	
QY	101	AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla	120	
Db	463	GATGTGCCACAGCGAGAGGGCCATTCCCGTGTCA	CGGAGGAGAGCCAGCTCGGCA	522
QY	121	ProSerSer	123	
Db	523	CCCTCGTCC	531	

### RESULT 3

RESOL 3  
US-10-316-253-110  
; Sequence 110, Application US/10316253  
; Publication No. US20030162706A1  
; GENERAL INFORMATION:  
; APPLICANT: The Procter & Gamble Company

```

; APPLICANT: Peters, Kevin
; APPLICANT: Thompson, Larry
; APPLICANT: Wang, Feng
; APPLICANT: Greis, Kenneth
; TITLE OF INVENTION: Angiogenesis Modulating Proteins
; FILE REFERENCE: 8865M
; CURRENT APPLICATION NUMBER: US/10/316,253
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: US 60/355,295
; PRIOR FILING DATE: 2002-02-08
; NUMBER OF SEQ ID NOS: 308
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 110
; LENGTH: 1271
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (159)..(749)
; OTHER INFORMATION:
; US-10-316-253-110

```

Alignment Scores:

Alignment Scores:		
Pred. No.:	5.05e-74	1271
Score:	577.50	Matches: 115
Percent Similarity:	82.19%	Conservative: 5
Best Local Similarity:	78.77%	Mismatches: 3
Query Match:	90.38%	Indels: 23
DB:	16	Gaps: 1

US-10-657-740-1 COPY 51 173 (1-123) x US-10-316-253-110 (1-1271)

QY	1	SerLeuPheArgThrValLeuAspSerGlyIleSerGlu	-----13
Db	309	TCTCTCTCCGACAGGTGGATCCGGCATCTCTGAGCTCATGCCATATGTGTTT	368
QY	14	-----ValArgSerAsp	17
Db	369	GTAATGCACCAACCATCATCTGTGGAAACCCAGAACACCCCGCAGGTCCCATCTGCAC	428
QY	18	ArgAspLysPheValIlePheLeuAspValIleHisPheSerProGluAspLeuThrVal	37
Db	429	CGGACAGAGTTGTTCATCTTCTGGATGTGAAGCACCTTCTCTCTGAGGACCTCACCGTG	488
QY	38	LysValGlnAspAspPheValGluIleHisGlyIleHisAsnGluArgGlnAspAspHis	57
Db	489	AAGGTACTGGGAAGATTCTGTGGAGATCCATGGCAACACAAACGAGAGCGAGGATGACCAT	548
QY	58	GlyTyrlleSerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSer	77
Db	549	GGCTACATTTCCCGTGAATTTACCGTCGTACCGTCTGCCTTCCAATGTGGACCACTCC	608
QY	78	AlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGln	97
Db	609	GCCTCTCTCTCTCTCTGTCTGCGGATGGCATGTGACCTTCTCTGSCCCCAAGGTCACG	668
QY	98	ThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysPro	117
Db	669	TCTGCTTGGATGTGGCCACAGCGAGAGGGCCATTTCCCGTGTCACGGNGGAGAACGCC	728
QY	118	ThrSerAlaProSerSer	123
Db	729	AGCTCGCACCTCGTCC	746

## RESULT 4

```

US-10-105-427-1
; Sequence 1, Application US/10105427
; Publication No. US20020177192A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific a
; TITLE OF INVENTION: Chimeric prote
; TITLE OF INVENTION: chaperrone-lik
; FILE REFERENCE: US- 649

```

; GENERAL INFORMATION:  
 ; APPLICANT: Council of Scientific and Industrial Research  
 ; TITLE OF INVENTION: Chimeric protein alpha BNAC crystallin with extraordinarily high  
 ; TITLE OF INVENTION: chaperone-like activity and a method thereof  
 ; FILE REFERENCE: US- 649

```
/ CURRENT APPLICATION NUMBER: US/10/105,427
/ CURRENT FILING DATE: 2002-07-03
/ NUMBER OF SEQ ID NOS: 2
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 1
/ LENGTH: 531
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: DNA sequence for chimeric alpha BNAC
US-10-105-427-1

Alignment Scores:
Pred. No.: 4.89e-73 Length: 531
Score: 567.00 Matches: 106
Percent Similarity: 97.41% Conservativeness: 7
Best Local Similarity: 91.38% Mismatches: 3
Query Match: 88.73% Indels: 0
DB: 13 Gaps: 0

US-10-657-740-1_COPY_51_173 (1-123) x US-10-105-427-1 (1-531)

QY 8 AspSerGlyLeuSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27
Db 184 GACACTGCTCTCAGAGATCGCGCTGGAGAGACAGAGTTCTCTGTCAACCTGGATGTG 243

QY 28 LysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHis 47
Db 244 AAGCACTTCTCCCGAGGACCTCACCGTGAGGTGAGGACGACACTTTGTGGAGATCCAC 303

QY 48 GlyLysHisGlnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArg 67
Db 304 GGAAGACACAAACGAGCGCCAGGACGACACCGGCTACATTTCCGTGAGTTCCACCGCGC 363

QY 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87
Db 364 TACGCGCTCGCTCCACACGGGACCAAGTGGCGCTCTCTGTGCTCCCTGTCTCCGATGC 423

QY 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107
Db 424 ATGCTGACCTTCTGTGGCCCAAGATCCAGACTGGCTGGATGCCACCCACCGCGGCGA 483

QY 108 AlaIleProValSerArgGluLysProThrSerAlaProSerSer 123
Db 484 GCCATCCCCGTGTCGGGGAGGAGAGACCCACCTCGGCTCCCTCGTCC 531

RESULT 5
US-10-029-386-26581
; Sequence 26581, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR C
; FILE REFERENCE: ABOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 26581
; LENGTH: 211
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR21.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.8
; OTHER INFORMATION: NT HIT: AF026952.1, EVALUE 0.00e+00
; FILE REFERENCE: ABOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 26581
; LENGTH: 211
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR21.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.8
; OTHER INFORMATION: SWISSPROT HIT: P02489, EVALUE 3.00e-36
; OTHER INFORMATION: EST_HUMAN HIT: BF726399.1, EVALUE 1.00e-113
```

```
/ OTHER INFORMATION: NT HIT: gi14780619, EVALUE 1.00e-115
US-10-029-386-26581
```

```
Alignment Scores:
Pred. No.: 3e-44 Length: 211
Score: 367.00 Matches: 70
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 57.43% Indels: 0
DB: 16 Gaps: 0
```

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-10-029-386-26581 (1-211)

```
QY 54 GlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeuProSerAsn 73
Db 1 CAGGACGACCAACGCGCTACATTTCCCGTGAGTTCCACCGCGCTACCGCTCGCTCCAAAC 60

QY 74 ValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGly 93
Db 61 GTGGACCAAGTCGGGCCCTCTCTGTCTCCCTGTCTGCGATGGCATGCTGACCTTCTGTGGC 120

QY 94 ProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProValSerArg 113
Db 121 CCCAAGATCCAGACTGGCTGGATGCCACCCACCGGAGGAGCCATCCCCGTGTGCGCG 180

QY 114 GluGluLysProThrSerAlaProSerSer 123
Db 181 GAGGAGAAGCCACCTCGGCTCCCTCGTCC 210
```

#### RESULT 6

```
US-10-029-386-12881
; Sequence 12881, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR C
; FILE REFERENCE: ABOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 12881
; LENGTH: 573
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR21.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.8
; OTHER INFORMATION: NT HIT: AF026952.1, EVALUE 0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: BF726856.1, EVALUE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P02489, EVALUE 1.00e-32
US-10-029-386-12881
```

```
Alignment Scores:
Pred. No.: 1.3e-43 Length: 573
Score: 367.00 Matches: 70
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 57.43% Indels: 0
DB: 16 Gaps: 0
```

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-10-029-386-12881 (1-573)

```
QY 54 GlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeuProSerAsn 73
Db 203 CAGGACGACCAACGCGCTACATTTCCCGTGAGTTCCACCGCGCTACCGCTCGCTCCAAAC 262
```

QY 74 ValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGly 93  
 Db 263 GTGGACCACTGGCGCTCTCTGCTCCCTGCTGCGGATGCTGACCTTCTGTGGC 322  
 QY 94 ProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProValSerArg 113  
 Db 323 CCCAAGATCCAGACTGGCTGATGCCACCCAGCCGAGGAGCCATCCCGGTGTGGCG 382  
 QY 114 GluGluLysProThrSerAlaProSerSer 123  
 Db 383 GAGGAGAGCCCACTGGCTCCCTCGTCC 412

## RESULT 7

US-10-152-319A-1574  
 ; Sequence 1574, Application US/10152319A  
 ; Publication No. US20040072160A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Mendrick, Donna  
 ; APPLICANT: Porter, Mark  
 ; APPLICANT: Johnson, Kory  
 ; APPLICANT: Higgs, Brandon  
 ; APPLICANT: Castle, Arthur  
 ; APPLICANT: Elashoff, Michael  
 ; TITLE OF INVENTION: Molecular Toxicology Modeling  
 ; FILE REFERENCE: 4921-5089-US  
 ; CURRENT APPLICATION NUMBER: US/10/152,319A  
 ; PRIOR FILING DATE: 2002-05-22  
 ; PRIOR APPLICATION NUMBER: US 60/292,335  
 ; PRIOR FILING DATE: 2001-05-22  
 ; PRIOR APPLICATION NUMBER: US 60/297,523  
 ; PRIOR FILING DATE: 2001-06-13  
 ; PRIOR APPLICATION NUMBER: US 60/298,925  
 ; PRIOR FILING DATE: 2001-06-19  
 ; PRIOR APPLICATION NUMBER: US 60/303,810  
 ; PRIOR FILING DATE: 2001-07-10  
 ; PRIOR APPLICATION NUMBER: US 60/303,807  
 ; PRIOR FILING DATE: 2001-07-10  
 ; PRIOR APPLICATION NUMBER: US 60/303,808  
 ; PRIOR FILING DATE: 2001-07-10  
 ; PRIOR APPLICATION NUMBER: US 60/315,047  
 ; PRIOR FILING DATE: 2001-08-28  
 ; PRIOR APPLICATION NUMBER: US 60/324,928  
 ; PRIOR FILING DATE: 2001-09-27  
 ; PRIOR APPLICATION NUMBER: US 60/330,867  
 ; PRIOR FILING DATE: 2001-11-01  
 ; PRIOR APPLICATION NUMBER: US 60/330,462  
 ; PRIOR FILING DATE: 2001-10-22  
 ; Remaining Prior Application data removed - See File Wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 2221  
 ; SOFTWARE: Patent In Ver. 2.1  
 ; SEQ ID NO 1574  
 ; LENGTH: 528  
 ; TYPE: DNA  
 ; ORGANISM: Rattus norvegicus  
 ; FEATURE:  
 ; OTHER INFORMATION: Genbank Accession No. NM\_012935  
 US-10-152-319A-1574

Alignment Scores:  
 Pred. No.: 8,76e-43 Length: 528  
 Score: 361.00 Matches: 67  
 Percent Similarity: 79.49% Conservative: 26  
 Best Local Similarity: 57.26% Mismatches: 18  
 Query Match: 56.49% Indels: 6  
 DB: 17 Gaps: 2

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-10-152-319A-1574 (1-528)

QY 7 LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26  
 Db 181 ATTGACACTGGGCTCTCAGAGATGCGTATGGAGAGGACAGGTTCTCTGTGAACCTGGAC 240  
 QY 27 ValLysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGluIle 46

Db 241 GTGAGACCACTTCTCTCCAGAGAACTCAAAGTCAAGGTTCTGGGAGACGTGATTGAGGTG 300  
 QY 47 HisGlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArg 66  
 Db 301 CACGCGAAGCAGGAGAGCGCAGCAGCAATGCTTCATCTCCAGGGAGTTCACAGG 360  
 QY 67 ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86  
 Db 361 AAGTACCGGATCCAGCGGACGTGATCTCTCACCATTTACTTCTCCCTGTTCATCGGAT 420  
 QY 87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106  
 Db 421 GGAGTCTCTCACTGTGAATGGACCAAGGAAACAG-----GCCTCTGGCCCTGAG 468  
 QY 107 ArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 121  
 Db 469 CGCACCATTCCTCCATCACCCGCTGAAGAGAGAGCCCTGCTGCTCAGTCGAGCCCT 519

## RESULT 8

US-10-486-706-455  
 ; Sequence 455, Application US/10486706  
 ; Publication No. US20050071088A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: LANDFIELD, PHILIP W.  
 ; APPLICANT: BLALOCK, ERIC M.  
 ; APPLICANT: CHEN, KUBY-CHU  
 ; APPLICANT: FOSTER, THOMAS C.  
 ; TITLE OF INVENTION: GENE EXPRESSION PROFILE BIOMARKERS AND THERAPEUTIC TARGETS FOR  
 ; TITLE OF INVENTION: BRAIN AGING AND AGE-RELATED COGNITIVE IMPAIRMENT  
 ; FILE REFERENCE: 50229-426  
 ; CURRENT APPLICATION NUMBER: US/10/486,706  
 ; CURRENT FILING DATE: 2004-02-13  
 ; PRIOR APPLICATION NUMBER: PCT/US02/25607  
 ; PRIOR FILING DATE: 2002-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/311,343  
 ; PRIOR FILING DATE: 2001-08-13  
 ; NUMBER OF SEQ ID NOS: 461  
 ; SOFTWARE: Patent In version 3.2  
 ; SEQ ID NO 455  
 ; LENGTH: 689  
 ; TYPE: DNA  
 ; ORGANISM: Rattus rattus  
 US-10-486-706-455

## Alignment Scores:

Pred. No.: 1.3e-42 Length: 689  
 Score: 261.00 Matches: 67  
 Percent Similarity: 79.49% Conservative: 26  
 Best Local Similarity: 57.26% Mismatches: 18  
 Query Match: 56.49% Indels: 6  
 DB: 19 Gaps: 2

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-10-486-706-455 (1-689)

QY 7 LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26  
 Db 193 ATTGACACTGGGCTCTCAGAGATGCGTATGGAGAGGACAGGTTCTCTGTGAACCTGGAC 252  
 QY 27 ValLysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGluIle 46  
 Db 253 GTGAGACCACTTCTCTCCAGAGAACTCAAAGTCAAGGTTCTGGGAGACGTGATTGAGGTG 312  
 QY 47 HisGlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArg 66  
 Db 313 CACGCGAAGCAGGAGAGCGCAGCAGCAATGCTTCATCTCCAGGGAGTTCACAGG 372  
 QY 67 ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86  
 Db 373 AAGTACCGGATCCAGCGGACGTGATCTCTCACCATTTACTTCTCCCTGTTCATCGGAT 432  
 QY 87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106



```
; Publication No. US20030148295A1
; GENERAL INFORMATION:
; APPLICANT: WAN, JACKSON
; APPLICANT: WANG, YIXIN
; TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE
; FILE REFERENCE: 15117.0012
; CURRENT APPLICATION NUMBER: US/10/101,510
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/276,947
; PRIOR FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 805
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 250
; LENGTH: 548
; TYPE: DNA
; ORGANISM: Oryctolagus cuniculus
US-10-101-510-250

Alignment Scores:
Pred. No.: 2,558-42 Length: 548
Score: 358.00 Matches: 67
Percent Similarity: 78.63% Conservative: 25
Best Local Similarity: 57.26% Mismatches: 19
Query Match: 56.03% Indels: 6
DB: 15 Gaps: 2

US-10-657-740-1_COPY_51_173 (1-123) x US-10-101-510-250 (1-548)

QY 7 LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26
DB 201 ATTGACACTGGACTCTCAGAGATGCGCTGGAGAGGACAGGTTCTGTCAACCTGGAT 260
QY 27 ValLysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGluLe 46
DB 261 GTGAAGCACTTCTCCCGAGAGAGCTCAAGGTCAAAGTGTGGTGATGTGATTGAGGTG 320
QY 47 HisGlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArg 66
DB 321 CACGCCAACATGAAGAGCGCCAGATGAACATGTTTCATCTCCAGGAGTTCCACAGG 380
QY 67 ArgTyrArgLeuProSerAsnValArgGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86
DB 381 AAATACCGATCCAGCTGATGTGACCTCTCACCATTACTTCATCTCTGTCATCTGAT 440
QY 87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106
DB 441 GGGGTCTCACTGTGAATGGACCAAGGAAGCAA-----GCCCTGGCCCGAG 488
QY 107 ArgAlaIleProValSerArgGluLysPro-----ThrSerAlaPro 121
DB 489 CGCACCATTCCTCAACCCGTGAAGAGAAGCCTGCTGTCACTGCAGCCCCC 539

RESULT 12
US-10-770-668-35
; Sequence 35, Application US/10770668
; Publication No. US20040191843A1
; GENERAL INFORMATION:
; APPLICANT: Wright, Susan C.
; APPLICANT: Larrick, James W.
; APPLICANT: Nock, Steffen R.
; APPLICANT: Wilson, David S.
; TITLE OF INVENTION: Cell-Killing Molecules and Methods of Use Thereof
; FILE REFERENCE: ABSALUS-08602
; CURRENT APPLICATION NUMBER: US/10/770,668
; CURRENT FILING DATE: 2004-02-02
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 35
; LENGTH: 528
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-770-668-35

Alignment Scores:
Pred. No.: 9,358-42 Length: 528
Score: 354.00 Matches: 66
Percent Similarity: 78.45% Conservative: 25
Best Local Similarity: 56.90% Mismatches: 19
Query Match: 55.40% Indels: 6
DB: 18 Gaps: 2

US-10-657-740-1_COPY_51_173 (1-123) x US-10-770-668-35 (1-528)

QY 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27
DB 184 GACACTGGACTCTCAGAGATGCGCTGGAGAGGACAGGTTCTGTCAACCTGGATGG 243
QY 28 LysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGluLeHis 47
DB 244 AAGCACTTCTCCCGAGAGAACTCAAAGTAAAGTGTGGGAGATGTGATTGAGGTGCAT 303
QY 48 GlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArgArg 67
DB 304 GGAACCAATGAAGAGCGCCAGATGAACATGTTTCATCTCCAGGAGTTCCACAGGAAA 363
QY 68 TyrArgLeuProSerAsnValArgGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87
DB 364 TACCGATCCAGCTGATGTAGACCTCTCACCATTACTTCATCTCTGTCATCTGATGG 423
QY 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107
DB 424 GTCTCTCACTGTGAATGGACCAAGGAACAG-----GTCTCTGGCCCTGAGCGC 471
QY 108 AlaIleProValSerArgGluLysPro-----ThrSerAlaPro 121
DB 472 ACCATTCCCATCCCGTGAAGAGAAGCCTGCTGTCAACCGAGCCCCC 519

RESULT 13
US-09-954-456-514
; Sequence 514, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 514
; LENGTH: 691
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-514
```

```
Alignment Scores:
Pred. No.: 1.39e-41 Length: 691
Score: 354.00 Matches: 66
Percent Similarity: 78.45% Conservative: 25
Best Local Similarity: 56.90% Mismatches: 19
Query Match: 55.40% Indels: 6
DB: 9 Gaps: 2

US-10-657-740-1_COPY_51_173 (1-123) x US-09-954-456-514 (1-691)

Qy 8 AspSerGlyLeuSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27
Db 209 GACACTGGACTCTCAGAGATGCGCTCGAAGAGACACAGGTTCTCTGTCAACCTGGATGG 268
Qy 28 LysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGluIleHis 47
Db 269 AAGCACTTCTCCCAAGAGGAGCACTAAAGTTAAGGTGTTGGGAGATGTGATTGAGGTGCAT 328
Qy 48 GlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArgArg 67
Db 329 GGAACATCATGAAGAGCGCCAGGATGAACATGGTTTTCATCTCCAGGGAGTTCACACAGAAA 388
Qy 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87
Db 389 TACCGGATCCAGCTGATGATAGACCTCTCACCATTACTTCTCCTCTCATCTGATGG 448
Qy 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107
Db 449 GTCTCTACTGTGAATGGACCAAGAAACAG-----GTCTCTGCCCTGAGCGC 496
Qy 108 AlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 121
Db 497 ACCATTCCCATCCCGTGAAGAGAGCGCTCTGTCTCACCAGCGCCCC 544

RESULT 14
US-09-960-706-869
; Sequence 869, Application US/09960706
; Publication No. US20030134280A1
; GENERAL INFORMATION:
; APPLICANT: Munger, William E.
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperplasia
; FILE REFERENCE: 44921-5029-01US
; CURRENT APPLICATION NUMBER: US/09/960,706
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: 60/223,323
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: 09/873,319
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 869
; LENGTH: 691
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20030134280A1 S45630
US-09-960-706-869

Alignment Scores:
Pred. No.: 1.39e-41 Length: 691
Score: 354.00 Matches: 66
Percent Similarity: 78.45% Conservative: 25
Best Local Similarity: 56.90% Mismatches: 19
Query Match: 55.40% Indels: 6
DB: 10 Gaps: 2

US-10-657-740-1_COPY_51_173 (1-123) x US-09-960-706-869 (1-691)

Qy 8 AspSerGlyLeuSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27
Db 209 GACACTGGACTCTCAGAGATGCGCTCGAAGAGACACAGGTTCTCTGTCAACCTGGATGG 268

US-09-960-706-869
; Sequence 869, Application US/09960706
; Publication No. US20030134280A1
; GENERAL INFORMATION:
; APPLICANT: Munger, William E.
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperplasia
; FILE REFERENCE: 44921-5029-01US
; CURRENT APPLICATION NUMBER: US/09/960,706
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: 60/223,323
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: 09/873,319
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 869
; LENGTH: 691
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20030134280A1 S45630
US-09-960-706-869

Alignment Scores:
Pred. No.: 1.39e-41 Length: 691
Score: 354.00 Matches: 66
Percent Similarity: 78.45% Conservative: 25
Best Local Similarity: 56.90% Mismatches: 19
Query Match: 55.40% Indels: 6
DB: 10 Gaps: 2

US-10-657-740-1_COPY_51_173 (1-123) x US-09-960-706-869 (1-691)

Qy 8 AspSerGlyLeuSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27
Db 209 GACACTGGACTCTCAGAGATGCGCTCGAAGAGACACAGGTTCTCTGTCAACCTGGATGG 268
```

```
Qy 28 LysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGluIleHis 47
Db 269 AAGCACTTCTCCCAAGAGGAGCACTAAAGTTAAGGTGTTGGGAGATGTGATTGAGGTGCAT 328
Qy 48 GlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArgArg 67
Db 329 GGAACATCATGAAGAGCGCCAGGATGAACATGGTTTTCATCTCCAGGGAGTTCACACAGAAA 388
Qy 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87
Db 389 TACCGGATCCAGCTGATGATAGACCTCTCACCATTACTTCTCCTCTCATCTGATGG 448
Qy 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107
Db 449 GTCTCTACTGTGAATGGACCAAGAAACAG-----GTCTCTGCCCTGAGCGC 496
Qy 108 AlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 121
Db 497 ACCATTCCCATCCCGTGAAGAGAGCGCTCTGTCTCACCAGCGCCCC 544

RESULT 15
US-09-873-319-566
; Sequence 566, Application US/09873319A
; Publication No. US20030134324A1
; GENERAL INFORMATION:
; APPLICANT: Munger, William E.
; APPLICANT: Kulkarni, Prakash
; APPLICANT: Getzenberg, Robert H.
; APPLICANT: Waga, Iwao
; APPLICANT: Yamamoto, Jun
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperplasia
; FILE REFERENCE: 44921-5029-US
; CURRENT APPLICATION NUMBER: US/09/873,319A
; CURRENT FILING DATE: 2001-06-05
; EARLIER APPLICATION NUMBER: US 60/223,323
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 755
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 566
; LENGTH: 691
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20030134324A1 S45630
US-09-873-319-566

Alignment Scores:
Pred. No.: 1.39e-41 Length: 691
Score: 354.00 Matches: 66
Percent Similarity: 78.45% Conservative: 25
Best Local Similarity: 56.90% Mismatches: 19
Query Match: 55.40% Indels: 6
DB: 10 Gaps: 2

US-10-657-740-1_COPY_51_173 (1-123) x US-09-873-319-566 (1-691)

Qy 8 AspSerGlyLeuSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27
Db 209 GACACTGGACTCTCAGAGATGCGCTCGAAGAGACACAGGTTCTCTGTCAACCTGGATGG 268
Qy 28 LysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGluIleHis 47
Db 269 AAGCACTTCTCCCAAGAGGAGCACTAAAGTTAAGGTGTTGGGAGATGTGATTGAGGTGCAT 328
Qy 48 GlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArgArg 67
Db 329 GGAACATCATGAAGAGCGCCAGGATGAACATGGTTTTCATCTCCAGGGAGTTCACACAGAAA 388
Qy 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87
Db 389 TACCGGATCCAGCTGATGATAGACCTCTCACCATTACTTCTCCTCTCATCTGATGG 448
```

```
Qy 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107
Db 449 GTCTCTCACTGTAATGACCAAGAAACAG-----GTCTTGGCCCTGAGGCGC 496

Qy 108 AlaIleProValSerArgGluLysPro-----ThrSerAlaPro 121
Db 497 ACCATTCCCATCACCCCGTGAAGAGAGCTGTCTGTACCGCAGCCCC 544

RESULT 16
US-10-657-740-1_COPY_51_173 (1-123) x US-09-873-367C-764 (1-691)
; Sequence 764, Application US/09873367C
; Publication No. US20030165839A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; APPLICANT: Soppet, Daniel
; APPLICANT: Endress, Gregory
; APPLICANT: Augustus, Meena
; APPLICANT: Ebner, Reinhard
; APPLICANT: Carter, Kenneth
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using
; FILE OF INVENTION: Signature Gene Sets
; FILE REFERENCE: 689290-64
; CURRENT APPLICATION NUMBER: US/09/873,367C
; PRIOR FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: U.S. 60/236,891
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: U.S. 60/236,842
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: U.S. 60/244,867
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: U.S. 60/245,084
; PRIOR FILING DATE: 2000-11-01
; NUMBER OF SEQ ID NOS: 1067
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 764
; LENGTH: 691
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-873-367C-764

Alignment Scores:
Pred. No.: 1,39e-41 Length: 691
Score: 354.00 Matches: 66
Percent Similarity: 78.45% Conservative: 25
Best Local Similarity: 56.90% Mismatches: 19
Query Match: 55.40% Indels: 6
DB: 10 Gaps: 2

US-10-657-740-1_COPY_51_173 (1-123) x US-09-873-367C-764 (1-691)
Qy 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27
Db 209 GACACTGGACTCTCAGAGATGGCTGGAGAGACAGAGTTCTCTGTCAACCTGGATGTG 268

Qy 28 LysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGluIleHis 47
Db 269 AAGCACTTCTCCCGAGAGAACTCAAAGTTAAGGTGTTGGAGATGTGATTGAGGTGCAT 328

Qy 48 GlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArg 67
Db 329 GGAATAACATGAAGAGCGCCAGATGAACATGTTTCATCTCCAGGGAGTTCCACAGGAAA 388

Qy 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87
Db 389 TACCGATCCAGCTGATGATAGACCTCTCACCATTACTTCTCATCTCTGATGATGG 448

Qy 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107
Db 449 GTCTCTCACTGTAATGACCAAGAAACAG-----GTCTTGGCCCTGAGGCGC 496

Qy 108 AlaIleProValSerArgGluLysPro-----ThrSerAlaPro 121
Db 497 ACCATTCCCATCACCCCGTGAAGAGAGCTGTCTGTACCGCAGCCCC 544

RESULT 17
US-10-133-937-61
; Sequence 61, Application US/10133937
; Publication No. US20030207278A1
; GENERAL INFORMATION:
; APPLICANT: Khan, Javed
; APPLICANT: Ringner, Markus
; APPLICANT: Peterson, Carsten
; APPLICANT: Meltzer, Paul
; TITLE OF INVENTION: METHODS FOR ANALYZING HIGH DIMENSIONAL DATA FOR CLASSIFYING,
; TITLE OF INVENTION: DIAGNOSING, PROGNOSTICATING, AND/OR PREDICTING DISEASES AND
; TITLE OF INVENTION: OTHER BIOLOGICAL STATES
; FILE REFERENCE: 11613.56US01
; CURRENT APPLICATION NUMBER: US/10/133,937
; CURRENT FILING DATE: 2002-11-04
; NUMBER OF SEQ ID NOS: 99
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 61
; LENGTH: 691
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-133-937-61

Alignment Scores:
Pred. No.: 1,39e-41 Length: 691
Score: 354.00 Matches: 66
Percent Similarity: 78.45% Conservative: 25
Best Local Similarity: 56.90% Mismatches: 19
Query Match: 55.40% Indels: 6
DB: 17 Gaps: 2

US-10-657-740-1_COPY_51_173 (1-123) x US-10-133-937-61 (1-691)
Qy 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27
Db 209 GACACTGGACTCTCAGAGATGGCTGGAGAGACAGAGTTCTCTGTCAACCTGGATGTG 268

Qy 28 LysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGluIleHis 47
Db 269 AAGCACTTCTCCCGAGAGAACTCAAAGTTAAGGTGTTGGAGATGTGATTGAGGTGCAT 328

Qy 48 GlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArg 67
Db 329 GGAATAACATGAAGAGCGCCAGATGAACATGTTTCATCTCCAGGGAGTTCCACAGGAAA 388

Qy 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87
Db 389 TACCGATCCAGCTGATGATAGACCTCTCACCATTACTTCTCATCTCTGATGATGG 448

Qy 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107
Db 449 GTCTCTCACTGTAATGACCAAGAAACAG-----GTCTTGGCCCTGAGGCGC 496

Qy 108 AlaIleProValSerArgGluLysPro-----ThrSerAlaPro 121
Db 497 ACCATTCCCATCACCCCGTGAAGAGAGCTGTCTGTACCGCAGCCCC 544

RESULT 18
US-10-172-118-686
; Sequence 686, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mac, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
```

```
; CURRENT APPLICATION NUMBER: US/10/172,118
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 686
; LENGTH: 691
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_001885
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-686

Alignment Scores:
Pred. No.: 1,39e-41 Length: 691
Score: 354.00 Matches: 66
Percent Similarity: 78.45% Conservative: 25
Best Local Similarity: 56.90% Mismatches: 19
Query Match: 55.40% Indels: 6
DB: 17 Gaps: 2

US-10-657-740-1_COPY_51_173 (1-123) x US-10-172-118-686 (1-691)

QY 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27
Db 209 GACATCGACTCTCAGAGATCGCGCTGGAGAGACAGAGTTCTCTGTCAACCTGGATGG 268
QY 28 LysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGluLeHis 47
Db 269 AAGCACTTCTCCCGAGGCGCTGAGATGAAAGTTAAGGTGTGGAGATGATGAGGTGCAT 328
QY 48 GlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArg 67
Db 329 GGAAACATGAAGAGCGCGAGGATGAACATGTTTCACTCCAGGAGTTCCACAGGAAA 388
QY 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87
Db 389 TACCGATCCAGCTGATGAGACCTCTCACCATTACTTCCCTGTCTGTCATCTGATGG 448
QY 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107
Db 449 GTCTCTCACTGTGAATGACCAAGAAACAG-----GTCTCTGGCCCTGAGCGC 496
QY 108 AlaIleProValSerArgGluLysPro-----ThrSerAlaPro 121
Db 497 ACCATTCCCATCCCGTGAAGAGAGCCTGTGTCCACCGAGCCCC 544

RESULT 19
US-10-159-563-61
; Sequence 61, Application US/10159563
; Publication No. US20040009154A1
; GENERAL INFORMATION:
; APPLICANT: Khan, Javed
; APPLICANT: Ringner, Markus
; APPLICANT: Peterson, Carsten
; APPLICANT: Meltzer, Paul
; TITLE OF INVENTION: SELECTIONS OF GENES AND METHODS OF USING THE SAME FOR
; FILE REFERENCE: 11613.56US11
; CURRENT APPLICATION NUMBER: US/10/159,563
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US 10/133,937
; PRIOR FILING DATE: 2002-04-25
; NUMBER OF SEQ ID NOS: 444
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61
; LENGTH: 691
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-159-563-61

Alignment Scores:
Pred. No.: 1,39e-41 Length: 691
Score: 354.00 Matches: 66
Percent Similarity: 78.45% Conservative: 25
Best Local Similarity: 56.90% Mismatches: 19
Query Match: 55.40% Indels: 6
DB: 17 Gaps: 2

US-10-657-740-1_COPY_51_173 (1-123) x US-10-342-887-686 (1-691)

QY 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27
Db 209 GACATCGACTCTCAGAGATCGCGCTGGAGAGACAGAGTTCTCTGTCAACCTGGATGG 268
QY 28 LysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGluLeHis 47
Db 269 AAGCACTTCTCCCGAGGCGCTGAGATGAAAGTTAAGGTGTGGAGATGATGAGGTGCAT 328
QY 48 GlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArg 67
Db 329 GGAAACATGAAGAGCGCGAGGATGAACATGTTTCACTCCAGGAGTTCCACAGGAAA 388
QY 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87
Db 389 TACCGATCCAGCTGATGAGACCTCTCACCATTACTTCCCTGTCTGTCATCTGATGG 448
QY 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107
Db 449 GTCTCTCACTGTGAATGACCAAGAAACAG-----GTCTCTGGCCCTGAGCGC 496
QY 108 AlaIleProValSerArgGluLysPro-----ThrSerAlaPro 121
Db 497 ACCATTCCCATCCCGTGAAGAGAGCCTGTGTCCACCGAGCCCC 544

RESULT 19
US-10-159-563-61
; Sequence 61, Application US/10159563
; Publication No. US20040009154A1
; GENERAL INFORMATION:
; APPLICANT: Khan, Javed
; APPLICANT: Ringner, Markus
; APPLICANT: Peterson, Carsten
; APPLICANT: Meltzer, Paul
; TITLE OF INVENTION: SELECTIONS OF GENES AND METHODS OF USING THE SAME FOR
; FILE REFERENCE: 11613.56US11
; CURRENT APPLICATION NUMBER: US/10/159,563
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US 10/133,937
; PRIOR FILING DATE: 2002-04-25
; NUMBER OF SEQ ID NOS: 444
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61
; LENGTH: 691
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-159-563-61

Alignment Scores:
Pred. No.: 1,39e-41 Length: 691
Score: 354.00 Matches: 66
Percent Similarity: 78.45% Conservative: 25
Best Local Similarity: 56.90% Mismatches: 19
Query Match: 55.40% Indels: 6
DB: 17 Gaps: 2

US-10-657-740-1_COPY_51_173 (1-123) x US-10-342-887-686 (1-691)

QY 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27
Db 209 GACATCGACTCTCAGAGATCGCGCTGGAGAGACAGAGTTCTCTGTCAACCTGGATGG 268
QY 28 LysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGluLeHis 47
Db 269 AAGCACTTCTCCCGAGGCGCTGAGATGAAAGTTAAGGTGTGGAGATGATGAGGTGCAT 328
QY 48 GlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArg 67
Db 329 GGAAACATGAAGAGCGCGAGGATGAACATGTTTCACTCCAGGAGTTCCACAGGAAA 388
QY 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87
Db 389 TACCGATCCAGCTGATGAGACCTCTCACCATTACTTCCCTGTCTGTCATCTGATGG 448
QY 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107
Db 449 GTCTCTCACTGTGAATGACCAAGAAACAG-----GTCTCTGGCCCTGAGCGC 496
QY 108 AlaIleProValSerArgGluLysPro-----ThrSerAlaPro 121
Db 497 ACCATTCCCATCCCGTGAAGAGAGCCTGTGTCCACCGAGCCCC 544

RESULT 20
US-10-342-887-686
; Sequence 686, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 686
; LENGTH: 691
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-686

Alignment Scores:
Pred. No.: 1,39e-41 Length: 691
Score: 354.00 Matches: 66
Percent Similarity: 78.45% Conservative: 25
Best Local Similarity: 56.90% Mismatches: 19
Query Match: 55.40% Indels: 6
DB: 17 Gaps: 2

US-10-657-740-1_COPY_51_173 (1-123) x US-10-342-887-686 (1-691)
```





US-10-657-740-1\_COPY\_51\_173 (1-123) x US-10-843-641A-3541 (1-691)

```
QY 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27
Db 209 GACACTGGACTCTCAGAGATGGCGCTGGAGAGGACAGGTTCTCTGCAACTGGATGTG 268
QY 28 LysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHis 47
Db 269 AAGCACTTCTCCAGAGAACTCAAGTTAAGTTGTTGGAGAGATGTTGATTGAGGTGCAT 328
QY 48 GlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArgArg 67
Db 329 GGAAGAACATGAAGAGCGCCAGGATGAACATGTTTTCATCTCCAGGAGTTCCACAGGAAA 388
QY 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87
Db 389 TACCGGATCCAGCTGATGTAGACCTCTCACCATTACTTCCCTGTCATCTGATGGG 448
QY 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107
Db 449 GTCTCTCACTGTGAATGGACCAAGAAACAG-----GTCTCTGGCCCTGAGCGC 496
QY 108 AlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 121
Db 497 ACCATTCCCATCACCCTGGAAGAGAGCTGCTGTCCAGCGAGCCCCC 544
```

## RESULT 23

US-10-367-057-187  
; Sequence 187, Application US/10367057  
; Publication No. US20050100554A1  
; GENERAL INFORMATION:  
; APPLICANT: Cuthill, Scott;  
; APPLICANT: Jackson, Amanda;  
; APPLICANT: Lewin, David A.;  
; APPLICANT: Ooi, Chean Eng  
; TITLE OF INVENTION: Complexes and Methods of Using Same  
; FILE REFERENCE: 21402-559  
; CURRENT APPLICATION NUMBER: US/10/367,057  
; CURRENT FILING DATE: 2003-02-14  
; PRIOR APPLICATION NUMBER: 60/256,911  
; PRIOR FILING DATE: 2002-02-14  
; NUMBER OF SEQ ID NOS: 198  
; SEQ ID NO 187  
; LENGTH: 691  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-367-057-187

Alignment Scores:  
Pred. No.: 1,39e-41 Length: 691  
Score: 354.00 Matches: 66  
Percent Similarity: 78.45% Conservative: 25  
Best Local Similarity: 56.90% Mismatches: 19  
Query Match: 55.40% Indels: 6  
DB: 19 Gaps: 2

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-10-367-057-187 (1-691)

```
QY 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27
Db 209 GACACTGGACTCTCAGAGATGGCGCTGGAGAGGACAGGTTCTCTGTCACCTGGATGTG 268
QY 28 LysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHis 47
Db 269 AAGCACTTCTCCAGAGAACTCAAGTTAAGTTGTTGGAGAGATGTTGATTGAGGTGCAT 328
QY 48 GlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArgArg 67
Db 329 GGAAGAACATGAAGAGCGCCAGGATGAACATGTTTTCATCTCCAGGAGTTCCACAGGAAA 388
QY 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 87
Db 497 ACCATTCCCATCACCCTGGAAGAGAGCTGCTGTCCAGCGAGCCCCC 544
```

```
Db 389 TACCGGATCCAGCTGATGTAGACCTCTCACCATTACTTTCATCTGTCATCTGATGGG 448
QY 88 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 107
Db 449 GTCTCTCACTGTGAATGGACCAAGAAACAG-----GTCTCTGGCCCTGAGCGC 496
QY 108 AlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 121
Db 497 ACCATTCCCATCACCCTGGAAGAGAGCTGCTGTCCAGCGAGCCCCC 544
```

## RESULT 24

US-09-960-352-4143  
; Sequence 4143, Application US/09960352  
; Patent No. US20020137139A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Nengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE REFERENCE: 16511.006/37-21(10298)C  
; CURRENT APPLICATION NUMBER: US/09/960,352  
; CURRENT FILING DATE: 2001-09-24  
; NUMBER OF SEQ ID NOS: 15112  
; SEQ ID NO 4143  
; LENGTH: 380  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; OTHER INFORMATION: Clone ID: 18-LIB3058-040-Q1-K1-E5  
US-09-960-352-4143

## Alignment Scores:

Pred. No.: 8.3e-39 Length: 380  
Score: 332.50 Matches: 65  
Percent Similarity: 75.65% Conservative: 22  
Best Local Similarity: 56.52% Mismatches: 23  
Query Match: 52.03% Indels: 5  
DB: 9 Gaps: 2

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-09-960-352-4143 (1-380)

```
QY 4 ArgThrValLeuAsp---SerGlyIleSerGluValArgSerAspArgAspLysPheVal 22
Db 12 CGCACCCAGCTGGATGCTCTGCCCCCTCAGAGATGGCGCTGGAGAGGACAGATTCTCT 71
QY 23 IlePheLeuAspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAsp 42
Db 72 GTCAACCTGGACCTGACGCACTTCTCCAGAGGAACTCAAGGCCAAGCTGCTGGAGAT 131
QY 43 PheValGluIleHisGlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArg 62
Db 132 GTGATTGAGTGCATGCGAAACATGAAGAGCGCCAGGATGAACATGTTTATCTCCCG 191
QY 63 GluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSer 82
Db 192 GAGTTCACAGGAAATACCGGATCCCGACTCCAGCTGAGTGGACCTCTGCCATTACTTATCC 251
QY 83 LeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAla 102
Db 252 CTGCTCTGTATGGGACCTCCTGATGATGCAAGCAAGGAAACAG-----GCC 299
QY 103 ThrHisAlaGluArgAlaIleProValSerArgGluGluLysPro 117
Db 300 TCCGGCCCTGAGCGCACCATTCCTCCATAACCCCTGAGAGAGAGCCG 344
```

## RESULT 25

US-09-960-352-4277  
; Sequence 4277, Application US/09960352  
; Patent No. US20020137139A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Nengbing

; APPLICANT: Byatt, John C.  
 ; APPLICANT: Mathialagan, Nagappan  
 ; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
 ; FILE REFERENCE: 16511.006/37-21(10298)C  
 ; CURRENT APPLICATION NUMBER: US/09/960,352  
 ; CURRENT FILING DATE: 2001-09-24  
 ; NUMBER OF SEQ ID NOS: 15112  
 ; SEQ ID NO 4277  
 ; LENGTH: 449  
 ; TYPE: DNA  
 ; ORGANISM: Bos taurus  
 ; OTHER INFORMATION: Clone ID: 19-BOVMS1-021-Q1-E1-E3  
 US-09-960-352-4277

Alignment Scores:  
 Pred. No.: 1.98e-31 Length: 449  
 Score: 283.00 Matches: 48  
 Percent Similarity: 86.25% Conservative: 21  
 Best Local Similarity: 60.00% Mismatches: 11  
 Query Match: 44.29% Indels: 0  
 DB: 9 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-09-960-352-4277 (1-449)

QY 7 LeuAspSerGlyIleSerGluValArgSerAspArgAspPheValIlePheLeuAsp 26  
 DB 210 ATTGACACTGGCGCTCTCAGAGATGCGTCTGGAGAGACAGATTCTCTCAACCTGGAT 269  
 QY 27 ValIleHisPheSerProGluAspLeuThrValIleValGlnAspPheValGluIle 46  
 DB 270 GTGAGACACTCTCCCGCAGAGAACTCAAGGTCAGGTCCTGGAGATGTGATTGAGGTG 329  
 QY 47 HisGlyLysHisAsnGluArgGlnAspAspHisGlyIleSerArgGluPheHisArg 66  
 DB 330 CATGCAACATGAAGAGCGCCAGGATGAACATGGTTTATCTCCGGGAGTCCACAGG 389  
 QY 67 ArgTyArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86  
 DB 390 AAATACCGGATCCGACTGACGTGGAACCTCTCGCCATTACTTCACTCCCTGCTGTGAT 449

RESULT 26

US-10-242-535A-26398  
 ; Sequence 26398, Application US/0242535A  
 ; Publication No. US20040013663A1  
 ; GENERAL INFORMATION:

; APPLICANT: ChondroGene Inc.  
 ; APPLICANT: Liaw, C.C.  
 ; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
 ; FILE REFERENCE: 4231/2005  
 ; CURRENT APPLICATION NUMBER: US/10/242,535A  
 ; PRIOR FILING DATE: 2002-09-12  
 ; PRIOR APPLICATION NUMBER: US 10/085,783  
 ; PRIOR FILING DATE: 2002-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/305,340  
 ; PRIOR FILING DATE: 2001-07-13  
 ; PRIOR APPLICATION NUMBER: US 60/275,017  
 ; PRIOR FILING DATE: 2001-03-12  
 ; PRIOR APPLICATION NUMBER: US 60/271,955  
 ; PRIOR FILING DATE: 2001-02-28  
 ; NUMBER OF SEQ ID NOS: 58994  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 26398  
 ; LENGTH: 373  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (24)..(24)  
 ; OTHER INFORMATION: n is a, c, g, or t  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (227)..(227)  
 ; OTHER INFORMATION: n is a, c, g, or t  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (227)..(227)

; OTHER INFORMATION: n is a, c, g, or t  
 US-10-242-535A-26398

Alignment Scores:  
 Pred. No.: 1.78e-31 Length: 373  
 Score: 282.50 Matches: 60  
 Percent Similarity: 71.30% Conservative: 22  
 Best Local Similarity: 52.17% Mismatches: 19  
 Query Match: 44.21% Indels: 15  
 DB: 17 Gaps: 4

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-10-242-535A-26398 (1-373)

QY 14 ValArgSerAspArgAspLysPheValIlePheLeuAspValIleHisPheSerProGlu 33  
 DB 3 ATGCGCTGGAGAGACAGAGNTCTCTCAACCTGGATGTGAAGCACTTCTCCCGAGAG 62  
 QY 34 AspLeuThrValIleValGlnAspPheValGluIleHisGlyLysHisAsnGluArg 53  
 DB 63 GAATCAAGTTAAGGTGTGGAGATGTGATGAGTGCATGGAAACATGAAGAGCGC 122  
 QY 54 GlnAspAspHisGlyIleSerArgGluPheHisArgTyArgLeuProSerAsn 73  
 DB 123 CAGGATGAACATGGTTTCACTCCAGGAGTTCACAGGAATACCGGATCCAGCTGAT 182  
 QY 74 ValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGly 93  
 DB 183 GTAGACCTCTCACTTACTTCACTCCCTGCTCATCTGATGGG-----GGN 227  
 QY 94 ProLysIle-----GlnThrGlyLeuAspAlaThrHisAlaGluArgAla 108  
 DB 228 CCTCACTGTGAATGGACCAAGGAAACAGGTCTCTG-GCCCTGAGCCCA-----CCA 280  
 QY 109 IleProValSerArgGluLysPro-----ThrSerAlaPro 121  
 DB 281 TTTCCTCATCCCGTGAGAGAGGCTGTGTCAACCGAGCCCC 325

RESULT 27

US-10-085-783A-26398  
 ; Sequence 26398, Application US/10085783A  
 ; Publication No. US20040037841A1  
 ; GENERAL INFORMATION:

; APPLICANT: ChondroGene Inc.  
 ; APPLICANT: Liaw, C.C.  
 ; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
 ; FILE REFERENCE: 4231/2002  
 ; CURRENT APPLICATION NUMBER: US/10/085,783A  
 ; PRIOR FILING DATE: 2002-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/305,340  
 ; PRIOR FILING DATE: 2001-07-13  
 ; PRIOR APPLICATION NUMBER: US 60/275,017  
 ; PRIOR FILING DATE: 2001-03-12  
 ; PRIOR APPLICATION NUMBER: US 60/271,955  
 ; PRIOR FILING DATE: 2001-02-28  
 ; NUMBER OF SEQ ID NOS: 58994  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 26398  
 ; LENGTH: 373  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (24)..(24)  
 ; OTHER INFORMATION: n is a, c, g, or t  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (227)..(227)  
 ; OTHER INFORMATION: n is a, c, g, or t  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (227)..(227)

Alignment Scores:  
 Pred. No.: 1.78e-31 Length: 373  
 Score: 282.50 Matches: 60



;; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL  
;; GENE EXPRESSION  
;; NUMBER OF SEQUENCES: 1508  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
;; STREET: 3174 PORTER DRIVE  
;; CITY: PALO ALTO  
;; STATE: CALIFORNIA  
;; COUNTRY: USA  
;; ZIP: 94304  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/10/641,643  
;; FILING DATE: 14-AUG-2003  
;; CLASSIFICATION: <Unknown>  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: <Unknown>  
;; FILING DATE: <Unknown>  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Zeller, Karen J.  
;; REGISTRATION NUMBER: 37,071  
;; REFERENCE/DOCKET NUMBER: PA-0001 US  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (650) 855-0555  
;; TELEFAX: (650) 845-4166  
;; INFORMATION FOR SEQ ID NO: 1300:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 789 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; IMMEDIATE SOURCE:  
;; LIBRARY: GENBANK  
;; CLONE: G32477  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 1300 :  
US-10-641-643-1300

Alignment Scores:  
Pred. No.: 6,51e-28 Length: 789  
Score: 261.50 Matches: 55  
Percent Similarity: 65.18% Conservative: 18  
Best Local Similarity: 49.11% Mismatches: 30  
Query Match: 40.92% Indels: 10  
DB: 17 Gaps: 1

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-10-641-643-1300 (1-789)

Qy 7 LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26  
Db 293 CTCAGCAGCGGGTCTCGAGATCGGCACACTGCGGACCGGTGCGCGGTGTCCTGGAT 352  
Qy 27 VallysHisPheSerProGluAspLeuThrVallysValGlnAspAspPheValGlu 46  
Db 353 GTCAACCACTTGGCCCGGAGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATC 412  
Qy 47 HisGlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArg 66  
Db 413 ACCGCAAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 472  
Qy 67 ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86  
Db 473 AAATACACGCTGCCCCCGGTGTGAGACCCCAAGTTTCCTCTCTCTCTCTCTCTCTCTGAG 532  
Qy 87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106  
Db 533 GGCACTGACCGGTGA-GGCCCC-----CATGCCCAA 564  
Qy 107 ArgAlaIleProValSerArgGluGluLysProThr 118  
::: ||| |||:::

Db 565 GCTAGCCAGCAGTCCAAACGAGATCACCATCCCACT 600

RESULT 31

US-10-153-668-283  
; Sequence 283, Application US/10153668  
; Publication No. US20030092616A1  
; GENERAL INFORMATION:  
; APPLICANT: HONDA, Goichi  
; APPLICANT: MATSUDA, Akio  
; APPLICANT: MURAMATSU, Shuji  
; APPLICANT: ISHIZAWA, Kenya  
; TITLE OF INVENTION: STAT6 Activating Gene  
; FILE REFERENCE: 1254-0207P  
; CURRENT APPLICATION NUMBER: US/10/153,668  
; PRIOR FILING DATE: 2002-05-24  
; PRIOR APPLICATION NUMBER: US 60/293,172  
; PRIOR FILING DATE: 2001-05-25  
; PRIOR APPLICATION NUMBER: US 60/316,031  
; PRIOR FILING DATE: 2001-08-31  
; PRIOR APPLICATION NUMBER: US 60/328,403  
; PRIOR FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: JP 2001-157043  
; PRIOR FILING DATE: 2001-05-25  
; PRIOR APPLICATION NUMBER: JP 2001-260681  
; PRIOR FILING DATE: 2001-08-30  
; PRIOR APPLICATION NUMBER: JP 2001-313175  
; PRIOR FILING DATE: 2001-10-10  
; NUMBER OF SEQ ID NOS: 488  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 283  
; LENGTH: 847  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (108)..(722)  
US-10-153-668-283

Alignment Scores:

Pred. No.: 7,22e-28 Length: 847  
Score: 261.50 Matches: 55  
Percent Similarity: 65.18% Conservative: 18  
Best Local Similarity: 49.11% Mismatches: 30  
Query Match: 40.92% Indels: 10  
DB: 14 Gaps: 1

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-10-153-668-283 (1-847)

Qy 7 LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26  
Db 348 CTCAGCAGCGGGTCTCGAGATCGGCACACTGCGGACCGGTGCGCGGTGTCCTGGAT 407  
Qy 27 VallysHisPheSerProGluAspLeuThrVallysValGlnAspAspPheValGlu 46  
Db 408 GTCAACCACTTGGCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATC 467  
Qy 47 HisGlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArg 66  
Db 468 ACCGCAAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 527  
Qy 67 ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86  
Db 528 AAATACACGCTGCCCCCGGTGTGAGACCCCAAGTTTCCTCTCTCTCTCTCTCTCTGAG 587  
Qy 87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106  
Db 588 GGCACTGACCGGTGA-GGCCCC-----CATGCCCAA 619  
Qy 107 ArgAlaIleProValSerArgGluGluLysProThr 118  
::: ||| |||:::  
Db 620 GCTAGCCAGCAGTCCAAACGAGATCACCATCCCACT 655  
RESULT 32

```

US-09-969-034-4480
; Sequence 4480, Application US/09969034
; Publication No. US20040110668A1
; GENERAL INFORMATION:
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Astle, Jon H.
; APPLICANT: Carroll, Eddie III
; APPLICANT: Catino, Theodore J.
; APPLICANT: Molino, Gary A.
; APPLICANT: Thiagalingam, Arunthathi
; APPLICANT: Lewis, Marcia E.
; TITLE OF INVENTION: Nucleic Acid Sequences Differentially
; FILE REFERENCE: 1657/1032
; CURRENT APPLICATION NUMBER: US/09/969,034
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 60/237,271
; NUMBER OF SEQ ID NOS: 4494
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4480
; LENGTH: 865
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-969-034-4480

Alignment Scores:
Pred No.: 7,45e-28 Length: 865
Score: 261.50 Matches: 55
Percent Similarity: 65.18% Conservative: 18
Best Local Similarity: 49.11% Mismatches: 30
Query Match: 40.92% Indels: 10
DB: 11 Gaps: 1

US-10-657-740-1_COPY_51_173 (1-123) x US-09-969-034-4480 (1-865)
Qy 7 LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26
Db 348 CTCAGCAGCGGGTCTCGGAGATCGGCACACTGCGGACCGCTGGCGGTGCTCTGGAT 407
Qy 27 ValLysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGluIle 46
Db 408 GTCAACCACTTCGCCCGGACGAGCTGACGGTCAAGACCAAGATGGCGTGTGGAGATC 467
Qy 47 HisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArg 66
Db 468 ACCGGCAAGCAGCAGGAGCGGACGACGATGCTACATCTCCCGTGTCTCCCTGAG 527
Qy 67 ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86
Db 528 AAATACACGCTGCCCGGACGAGCGGACGACGATGCTACATCTCCCTGCTGCTCCCTGAG 587
Qy 87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106
Db 588 GGCACACTGACCGTGA-GGCCCC-----CATGCCCAA 619

RESULT 33
US-10-172-118-626
; Sequence 626, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
US-10-172-118-626
; Sequence 626, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118

```

```
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 626
; LENGTH: 865
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-626

Alignment Scores:
Pred. No.: 7,45e-28 Length: 865
Score: 261.50 Matches: 55
Percent Similarity: 65.18% Conservative: 18
Best Local Similarity: 49.11% Mismatches: 30
Query Match: 40.92% Indels: 10
DB: 17 Gaps: 1

US-10-657-740-1_COPY_51_173 (1-123) x US-10-342-887-626 (1-865)

Qy 7 LeuAspSerGlyIleSerProGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26
Db 348 CTCACGACGGGGTCTCGGAGATCCGGCACACTGCGGACCGCTGCGCGGTGTCCTGGAT 407
Qy 27 VallysHisPheSerProGluAspLeuThrVallysValGlnAspAspPheValGluIle 46
Db 408 GTCAACCACTTCGCCCGGACGAGCTGACGTCAAGACCAAGATGCTCCCGTGTTCACGCGG 467
Qy 47 HisGlyLysHisAsnGluArgGlnAspAspHisGlyIleSerArgGluPheHisArg 66
Db 468 ACCGCAAGCAGCAGGAGCGGCGAGCAGCAGTGTGACACCCCAAGTTTCCTCCTGTCCTGGAT 527
Qy 67 ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86
Db 528 AAATACACGCTGCCCGCGGTTGGACACCCCAAGTTTCCTCCTGTCCTGGAT 587
Qy 87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106
Db 588 GGCACACTGACGCTGA-GGCCCC-----CATGCCCAA 619
Qy 107 ArgAlaIleProValSerArgGluGluLysProThr 118
Db 620 GCTAGCCACGAGTCCACGAGATCACCATCCCACT 655

RESULT 35
US-09-880-107-3865
; Sequence 3865, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3865
; LENGTH: 1231
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 Z23090
US-09-880-107-3865

Alignment Scores:
Pred. No.: 1.25e-27 Length: 1231
Score: 261.50 Matches: 55
Percent Similarity: 65.18% Conservative: 18

US-10-657-740-1_COPY_51_173 (1-123) x US-10-342-887-626 (1-865)

Qy 7 LeuAspSerGlyIleSerProGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26
Db 1003 CTCACGACGGGGTCTCGGAGATCCGGCACACTGCGGACCGCTGCGCGGTGTCCTGGAT 1062
Qy 27 VallysHisPheSerProGluAspLeuThrVallysValGlnAspAspPheValGluIle 46
Db 1063 GTCAACCACTTCGCCCGGACGAGCTGACGTCAAGACCAAGATGCGGTGTCCTGGAT 1122
Qy 47 HisGlyLysHisAsnGluArgGlnAspAspHisGlyIleSerArgGluPheHisArg 66
```

```
Best Local Similarity: 49.11% Mismatches: 30
Query Match: 40.92% Indels: 10
DB: 9 Gaps: 1

US-10-657-740-1_COPY_51_173 (1-123) x US-09-880-107-3865 (1-1231)

Qy 7 LeuAspSerGlyIleSerProGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26
Db 732 CTCACGACGGGGTCTCGGAGATCCGGCACACTGCGGACCGCTGCGCGGTGTCCTGGAT 791
Qy 27 VallysHisPheSerProGluAspLeuThrVallysValGlnAspAspPheValGluIle 46
Db 792 GTCAACCACTTCGCCCGGACGAGCTGACGTCAAGACCAAGATGCGGTGTCCTGGAT 851
Qy 47 HisGlyLysHisAsnGluArgGlnAspAspHisGlyIleSerArgGluPheHisArg 66
Db 852 ACCGCAAGCAGCAGGAGCGGCGAGCAGCAGTGTGACACCCCAAGTTTCCTCCTGTCCTGGAT 911
Qy 67 ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86
Db 912 AAATACACGCTGCCCGCGGTTGGACACCCCAAGTTTCCTCCTGTCCTGGAT 971
Qy 87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106
Db 972 GGCACACTGACGCTGA-GGCCCC-----CATGCCCAA 1003
Qy 107 ArgAlaIleProValSerArgGluGluLysProThr 118
Db 1004 GCTAGCCACGAGTCCACGAGATCACCATCCCACT 1039

RESULT 36
US-10-100-957A-169
; Sequence 169, Application US/10100957A
; Publication No. US20030096322A1
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1A
; CURRENT APPLICATION NUMBER: US/10/100,957A
; CURRENT FILING DATE: 2002-03-19
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 169
; LENGTH: 1380
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GFP-HSP27
; NAME/KEY: CDS
; LOCATION: (1)..(1380)
US-10-100-957A-169

Alignment Scores:
Pred. No.: 1.48e-27 Length: 1380
Score: 261.50 Matches: 55
Percent Similarity: 65.18% Conservative: 18
Best Local Similarity: 49.11% Mismatches: 30
Query Match: 40.92% Indels: 10
DB: 14 Gaps: 1

US-10-657-740-1_COPY_51_173 (1-123) x US-10-100-957A-169 (1-1380)

Qy 7 LeuAspSerGlyIleSerProGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26
Db 1003 CTCACGACGGGGTCTCGGAGATCCGGCACACTGCGGACCGCTGCGCGGTGTCCTGGAT 1062
Qy 27 VallysHisPheSerProGluAspLeuThrVallysValGlnAspAspPheValGluIle 46
Db 1063 GTCAACCACTTCGCCCGGACGAGCTGACGTCAAGACCAAGATGCGGTGTCCTGGAT 1122
Qy 47 HisGlyLysHisAsnGluArgGlnAspAspHisGlyIleSerArgGluPheHisArg 66
```

Db 1123 ACCGCAAGCAGCAGGAGCGGAGGAGCTACATCTCCCGTGTTCACCGG 1182  
 QY 67 ArgTyArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86  
 Db 1183 AAATACAGCTGCCCGCCCGGTGTGGACCCCAAGTTTCTCTCTCCCTGTCCCTGAG 1242  
 QY 87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106  
 Db 1243 GGCACACTGACCGTGA-GGCCCC-----CATGCCCAA 1274  
 QY 107 ArgAlaIleProValSerArgGluGluLysProThr 118  
 Db 1275 GCTAGCCACGAGTCCCAACGAGATCACCATCCCACT 1310

## RESULT 37

US-09-917-800A-981/c  
 ; Sequence 981, Application US/09917800A  
 ; Patent No. US20020119462A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Mendrick, Donna  
 ; APPLICANT: Porter, Mark  
 ; APPLICANT: Johnson, Kory  
 ; APPLICANT: Castle, Arthur  
 ; APPLICANT: Elashoff, Michael  
 ; APPLICANT: Gene Logic, Inc.  
 ; TITLE OF INVENTION: Molecular Toxicology Modeling  
 ; FILE REFERENCE: 44921-5038-US  
 ; CURRENT APPLICATION NUMBER: US/09/917,800A  
 ; CURRENT FILING DATE: 2001-07-31  
 ; PRIOR APPLICATION NUMBER: US 60/222,040  
 ; PRIOR FILING DATE: 2000-07-31  
 ; PRIOR APPLICATION NUMBER: US 60/222,880  
 ; PRIOR FILING DATE: 2000-11-02  
 ; PRIOR APPLICATION NUMBER: US 60/290,029  
 ; PRIOR FILING DATE: 2001-05-11  
 ; PRIOR APPLICATION NUMBER: US 60/290,645  
 ; PRIOR FILING DATE: 2001-05-15  
 ; PRIOR APPLICATION NUMBER: US 60/292,336  
 ; PRIOR FILING DATE: 2001-05-22  
 ; PRIOR APPLICATION NUMBER: US 60/295,798  
 ; PRIOR FILING DATE: 2001-06-06  
 ; PRIOR APPLICATION NUMBER: US 60/297,457  
 ; PRIOR FILING DATE: 2001-06-13  
 ; PRIOR APPLICATION NUMBER: US 60/298,884  
 ; PRIOR FILING DATE: 2001-06-19  
 ; PRIOR APPLICATION NUMBER: US 60/303,459  
 ; PRIOR FILING DATE: 2001-07-09  
 ; NUMBER OF SEQ ID NOS: 1740  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 981  
 ; LENGTH: 604  
 ; TYPE: DNA  
 ; ORGANISM: Rattus norvegicus  
 ; FEATURE:

US-09-917-800A-981  
 ; OTHER INFORMATION: Genbank Accession No. US20020119462A1 A1176658

Alignment Scores:  
 Pred. No.: 5,2e-28 Length: 604  
 Score: 261.00 Matches: 57  
 Percent Similarity: 65.18% Conservative: 16  
 Best Local Similarity: 50.89% Mismatches: 33  
 Query Match: 40.85% Indels: 6  
 DB: 9 Gaps: 2

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-09-917-800A-981 (1-604)

QY 7 LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26  
 Db 492 CTCAGTAGCGGTGCTCAGAGATCCGACAGCGCCGATGCTGGCCGCTGCTCCCTGGAC 433  
 QY 27 VallysHisPheSerProGluAspLeuThrVallysValGlnAspAspPheValGluIle 46

Db 432 GTCAACCACTTCGCTCCTGAGGAGCTCACAGTTAAGACCAAGGAGCGTGTGGAGATC 373  
 QY 47 HisGlyLysHisAsnGluArgGlnAspHisGlyTyrlleSerArgGluPheHisArg 66  
 Db 372 ACTGGCAAGCAGCAAGAAGGAGGATGAACATGGCTACATCTCTCGGTGCTTCACCGG 313  
 QY 67 ArgTyArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86  
 Db 312 AAATACAGCTCCCTCCAGGTGTGGACCCCACTTGGTGTCTCTTCTCTCTCCCTGAG 253  
 QY 87 GlyMetLeuThrPheCysGly-----ProLysIleGlnThrGlyLeuAspAlaThrHis 104  
 Db 252 GGCACACTCACGGTGAAGGCTCCGTCGCCAAAGCAGTCA-----CAATCA 205  
 QY 105 AlaGluArgAlaIleProValSerArgGluGluLys 116  
 Db 204 GCGGAGATCACCATTCCGTCACCTTTCAGGCCCGT 169

## RESULT 38

US-10-060-036-43  
 ; Sequence 43, Application US/10060036  
 ; Publication No. US20030073144A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Benson, Darin R.  
 ; APPLICANT: Kalos, Michael D.  
 ; APPLICANT: Lodes, Michael J.  
 ; APPLICANT: Persing, David H.  
 ; APPLICANT: Hepler, William T.  
 ; APPLICANT: Jjiang, Yuguu

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
 ; FILE REFERENCE: 210121.566  
 ; CURRENT APPLICATION NUMBER: US/10/060,036  
 ; CURRENT FILING DATE: 2002-01-30  
 ; NUMBER OF SEQ ID NOS: 4560  
 ; SOFTWARE: Fast-Seq for Windows Version 4.0  
 ; SEQ ID NO 43  
 ; LENGTH: 599  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-10-060-036-43

Alignment Scores:  
 Pred. No.: 7,21e-28 Length: 599  
 Score: 260.00 Matches: 55  
 Percent Similarity: 65.18% Conservative: 18  
 Best Local Similarity: 49.11% Mismatches: 30  
 Query Match: 40.69% Indels: 10  
 DB: 14 Gaps: 1

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-10-060-036-43 (1-599)

QY 7 LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26  
 Db 254 CTCAGAGCGGGTCTCGAGATCCGGCACACCTCGGACCGGTGGCGGTGCTCCCTGGAT 313

QY 27 VallysHisPheSerProGluAspLeuThrVallysValGlnAspAspPheValGluIle 46  
 Db 314 GTCAACCACTTCGCCCCGAGCGCTGACGGTCAAGACCAAGATGGCGTGTGGAGATC 373

QY 47 HisGlyLysHisAsnGluArgGlnAspHisGlyTyrlleSerArgGluPheHisArg 66  
 Db 374 ACCGCAAGCAGCAGGAGCGGACGAGCATGCTACATCTCCCGTGTTCACGCGG 433

QY 67 ArgTyArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86  
 Db 434 AAATACAGCTGCCCGCCCGGTGTGGACCCCAAGTTTCTCTCTCTCTCCCTGAG 493

QY 87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106  
 Db 494 GGCACACTGACCGTGA-GGCCCC-----CATGCCCAA 525



```
Qy 107 ArgAlaIleProValSerArgGluGluLysProThr 118
Db 526 GCTAGCCACGCGAGTCCACGAGATCACCATCCCACT 561
RESULT 39
US-10-764-420-43
; Sequence 43, Application US/10764420
; Publication No. US20050084872A1
; GENERAL INFORMATION:
; APPLICANT: Lum, Pek Yee
; APPLICANT: Tan, YeJun
; APPLICANT: Dai, Hongyue
; TITLE OF INVENTION: Methods For Determining Whether An Agent
; FILE OF INVENTION: Possesses A Defined Biological Activity
; FILE REFERENCE: ROSA122057
; CURRENT APPLICATION NUMBER: US/10/764,420
; CURRENT FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US 60/442,797
; PRIOR FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: US 60/474,413
; PRIOR FILING DATE: 2003-05-30
; NUMBER OF SEQ ID NOS: 3683
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 634
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-764-420-43
Alignment Scores:
Pred. No.: 7,84e-28 Length: 634
Score: 260.00 Matches: 57
Percent Similarity: 65.18% Conservative: 16
Best Local Similarity: 50.89% Mismatches: 33
Query Match: 40.69% Indels: 6
DB: 19 Gaps: 2
US-10-657-740-1_COPY_51_173 (1-123) x US-10-764-420-43 (1-634)
Qy 7 LeuAspSerGlyIleSerGluValArgSerArgAspLysPheValIlePheLeuAsp 26
Db 250 CTCAGCAGCGGGTCTCGAGATCCGACAGACGCGTGTGCGCGGTCTCCCTGAC 309
Qy 27 VallysHisPheSerProGluAspLeuThrVallysValGlnAspAspPheValGluLe 46
Db 310 GTCAACCACTTCGCTCCGAGAGCTCACAGTGAAGACCAAGGAGGCGTGTGAGATC 369
Qy 47 HisGlyLysHisAsnGluArgGlnAspAspHisGlyIleSerArgGluPheHisArg 66
Db 370 ACTGCAAGCAGCAAGAAAGGAGGACGACCAATGCTACATCTCTCGGTGCTTCACCCGG 429
Qy 67 ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86
Db 430 AAATACACGCTCCCTCCAGGTGTGACCCACCCCTAGTGTCTCTTCCTTATCCCTGAG 489
Qy 87 GlyMetLeuThrPheCysGly-----ProLysIleGlnThrGlyLeuAspAlaThrHis 104
Db 490 GGCACACTTACCGTGGAGGCTCCGTTGCCAAAGCAGTCACG-----CAGTCA 537
Qy 105 AlaGluArgAlaIleProValSerArgGluGluLys 116
Db 538 GCGGAGATCACCATTCCGGTTACTTTTCGAGGCCCGG 573
RESULT 40
US-09-917-800A-1428
; Sequence 1428, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Cascle, Arthur
; APPLICANT: Elashoff, Michael
```

```
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1428
; LENGTH: 787
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 M86389
US-09-917-800A-1428
Alignment Scores:
Pred. No.: 1.08e-27 Length: 787
Score: 260.00 Matches: 57
Percent Similarity: 65.18% Conservative: 16
Best Local Similarity: 50.89% Mismatches: 33
Query Match: 40.69% Indels: 6
DB: 9 Gaps: 2
US-10-657-740-1_COPY_51_173 (1-123) x US-09-917-800A-1428 (1-787)
Qy 7 LeuAspSerGlyIleSerGluValArgSerArgAspLysPheValIlePheLeuAsp 26
Db 296 CTCAGCAGCGGTGTCTCAGAGATCCGACAGACGCGCGATCGTGGCGGTGTCCCTGAC 355
Qy 27 VallysHisPheSerProGluAspLeuThrVallysValGlnAspAspPheValGluLe 46
Db 356 GTCAACCACTTCGCTCCCTGAGGAGCTCACAGTTAAGACCAAGGAGGCGTGTGAGATC 415
Qy 47 HisGlyLysHisAsnGluArgGlnAspAspHisGlyIleSerArgGluPheHisArg 66
Db 416 ACTGGCAAGCAGCAAGAAAGGAGGACGAGTGAATGCTACATCTCTCGGTGCTTCACCCGG 475
Qy 67 ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86
Db 476 AAATACACGCTCCCTCCAGGTGTGACCCACCCCTGGTGTCTCTTCCTTCCCTGAG 535
Qy 87 GlyMetLeuThrPheCysGly-----ProLysIleGlnThrGlyLeuAspAlaThrHis 104
Db 536 GGCACACTCACCCTGGAGGCTCCGCTGCCAAAGCAGTCACA-----CAATCA 583
Qy 105 AlaGluArgAlaIleProValSerArgGluGluLys 116
Db 584 GCGGAGATCACCATTCCGGTCACTTTTCGAGGCCCGT 619
RESULT 41
US-10-191-803-228
; Sequence 228, Application US/10191803
; Publication No. US20040014040A1
; GENERAL INFORMATION:
; APPLICANT: MENDRICK, Donna
```

APPLICANT: PORTER, Mark  
APPLICANT: JOHNSON, Kory  
APPLICANT: HIGGS, Brandon  
APPLICANT: CASTLE, Arthur  
APPLICANT: ELASHOFF, Michael  
TITLE OF INVENTION: Cardiotoxin Molecular Toxicology Modeling  
FILE REFERENCE: 44921-50900S  
CURRENT APPLICATION NUMBER: US/10/191,803  
CURRENT FILING DATE: 2002-07-10  
PRIOR APPLICATION NUMBER: US 60/303,819  
PRIOR FILING DATE: 2001-07-10  
PRIOR APPLICATION NUMBER: US 60/305,623  
PRIOR FILING DATE: 2001-07-17  
PRIOR APPLICATION NUMBER: US 60/369,351  
PRIOR FILING DATE: 2002-04-03  
PRIOR APPLICATION NUMBER: US 60/377,611  
PRIOR FILING DATE: 2002-05-06  
NUMBER OF SEQ ID NOS: 1140  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 228  
LENGTH: 787  
TYPE: DNA  
ORGANISM: Rattus norvegicus  
FEATURE:  
OTHER INFORMATION: Genbank Accession No. US20040014040A1 NM\_031970  
US-10-191-803-228

Alignment Scores:  
Pred. No.: 1.08e-27 Length: 787  
Score: 260.00 Matches: 57  
Percent Similarity: 65.18% Conservative: 16  
Best Local Similarity: 50.89% Mismatches: 33  
Query Match: 40.69% Indels: 6  
DB: 17 Gaps: 2

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-10-191-803-228 (1-787)

QY 7 LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26  
Db 314 CTGAGACGGGTGTGTGAGATCCGACAGCGCGGATCGTGGCGGTGCTCCCTGGAC 373  
QY 27 ValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGluLe 46  
Db 374 GTCAACCACTTCGCTCTGAGAGCTCAGATGAAAGACCAAGAGCGGTGGTGAGATC 433  
QY 47 HisGlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArg 66  
Db 434 ACTGGCAAGCAGCAAGAAAGGAGGATGAAATGCTATCTCTCGGTGCTTCAACCGG 493  
QY 67 ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86  
Db 494 AAATACACGCTCCCTCCAGGTGTGACCCACCTTGGTGTCTCTTCCCTGCTCCCTGAG 553  
QY 87 GlyMetLeuThrPheCysGly-----ProLysIleGlnThrGlyLeuAspAlaThrHis 104  
Db 554 GGCACACTCAGCTGAGGCTCCGCTGCCCAAGCAGTCACA-----CAATCA 601  
QY 105 AlaGluArgAlaIleProValSerArgGluGluLys 116  
Db 602 GCGGAGATCACCATTCCGGTCACTTTCGAGGCCCGT 637

## RESULT 42

US-10-152-319A-1963  
Sequence 1963, Application US/10152319A  
Publication No. US20040072160A1  
GENERAL INFORMATION:  
APPLICANT: Mendrick, Donna  
APPLICANT: Porter, Mark  
APPLICANT: Johnson, Kory  
APPLICANT: Higgs, Brandon  
APPLICANT: Castle, Arthur  
APPLICANT: Elashoff, Michael  
TITLE OF INVENTION: Molecular Toxicology Modeling

FILE REFERENCE: 44921-5089-US  
CURRENT APPLICATION NUMBER: US/10/152,319A  
CURRENT FILING DATE: 2002-05-22  
PRIOR APPLICATION NUMBER: US 60/292,335  
PRIOR FILING DATE: 2001-05-22  
PRIOR APPLICATION NUMBER: US 60/297,523  
PRIOR FILING DATE: 2001-06-13  
PRIOR APPLICATION NUMBER: US 60/298,925  
PRIOR FILING DATE: 2001-06-19  
PRIOR APPLICATION NUMBER: US 60/303,810  
PRIOR FILING DATE: 2001-07-10  
PRIOR APPLICATION NUMBER: US 60/303,807  
PRIOR FILING DATE: 2001-07-10  
PRIOR APPLICATION NUMBER: US 60/303,808  
PRIOR FILING DATE: 2001-07-10  
PRIOR APPLICATION NUMBER: US 60/315,047  
PRIOR FILING DATE: 2001-08-28  
PRIOR APPLICATION NUMBER: US 60/324,928  
PRIOR FILING DATE: 2001-09-27  
PRIOR APPLICATION NUMBER: US 60/330,867  
PRIOR FILING DATE: 2001-11-01  
PRIOR APPLICATION NUMBER: US 60/330,462  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 2221  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 1963  
LENGTH: 787  
TYPE: DNA  
ORGANISM: Rattus norvegicus  
FEATURE:  
OTHER INFORMATION: Genbank Accession No. NM\_031970  
US-10-152-319A-1963

## Alignment Scores:

Pred. No.: 1.08e-27 Length: 787  
Score: 260.00 Matches: 57  
Percent Similarity: 65.18% Conservative: 16  
Best Local Similarity: 50.89% Mismatches: 33  
Query Match: 40.69% Indels: 6  
DB: 17 Gaps: 2

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-10-152-319A-1963 (1-787)

QY 7 LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26  
Db 314 CTGAGACGGGTGTGTGAGATCCGACAGCGCGGATCGTGGCGGTGCTCCCTGGAC 373  
QY 27 ValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGluLe 46  
Db 374 GTCAACCACTTCGCTCTGAGAGCTCAGATGAAAGACCAAGAGCGGTGGTGAGATC 433  
QY 47 HisGlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArg 66  
Db 434 ACTGGCAAGCAGCAAGAAAGGAGGATGAAATGCTATCTCTCGGTGCTTCAACCGG 493  
QY 67 ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86  
Db 494 AAATACACGCTCCCTCCAGGTGTGACCCACCTTGGTGTCTCTTCCCTGCTCCCTGAG 553  
QY 87 GlyMetLeuThrPheCysGly-----ProLysIleGlnThrGlyLeuAspAlaThrHis 104  
Db 554 GGCACACTCAGCTGAGGCTCCGCTGCCCAAGCAGTCACA-----CAATCA 601  
QY 105 AlaGluArgAlaIleProValSerArgGluGluLys 116  
Db 602 GCGGAGATCACCATTCCGGTCACTTTCGAGGCCCGT 637

## RESULT 43

US-09-960-352-12622  
Sequence 12622, Application US/09960352  
Patent No. US20020137139A1  
GENERAL INFORMATION:

```

US-10-657-740-1_COPY_51_173 (1-123) x US-10-425-115-63403 (1-393)
QY 7 LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26
Db 9 CTNAGCAGCGGGTCTCGAGATCCGGACACTCGGCACCCTGCCGGGTGTCCTGGAT 68
QY 27 ValylsHisPheSerProGluAspLeuThrVallysValGlnAspAspPheValGluIle 46
Db 69 GTCAACACATTGCCCCGGACAGTGACGGTCAAGACCAAGGATGGCGTGTGGAGATC 128
QY 47 HisGlyIysHisAsnGluArgGlnAspAspHisGlyTyrlleSerArgGluPheHisArg 66
Db 129 ACCGGCTAGCACGAGGACGCGACGAGCATGGCTACATCTCCCGTGTCTACGGCG 188
QY 67 ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86
Db 189 AATACACGCTGCCCGCGGTGTGGACCCACCCAAAGTTTCTCTCTCTGTCCTCGAG 248
QY 87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisalaGlu 106
Db 249 GGCACACTGACCGTGA-GGCCCC-----CATGCCCAA 280
QY 107 ArgAlaIleProValSerArgGluGluLysProThr 118
Db 281 GCTAGCCACGCGAGTCCAACGAGATCACCATCCCAGT 316

RESULT 45
US-10-226-956-320
; Sequence 320, Application US/10226956
; Publication No. US20030060399A1
; GENERAL INFORMATION:
; APPLICANT: Brophy, Colleen
; APPLICANT: Komalavilas, Padmini
; APPLICANT: Panitch, Alyssa
; APPLICANT: Joshi, Lokesh
; APPLICANT: Seal, Brandon L.
; TITLE OF INVENTION: REAGENTS AND METHODS FOR SMOOTH MUSCLE THERAPIES
; FILE REFERENCE: ASU-1061-US
; CURRENT APPLICATION NUMBER: US/10/226,956
; CURRENT FILING DATE: 2002-08-23
; PRIOR APPLICATION NUMBER: 60/314,535
; PRIOR FILING DATE: 2001-08-23
; NUMBER OF SEQ ID NOS: 320
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 320
; LENGTH: 480
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Sequence encoding human HSP20
US-10-226-956-320

Alignment Scores:
Pred. No.: 1,29e-26 Length: 480
Score: 250.50 Matches: 46
Percent Similarity: 63.64% Conservative: 24
Best Local Similarity: 41.82% Mismatches: 25
Query Match: 39.20% Indels: 15
DB: 14 Gaps: 1

US-10-657-740-1_COPY_51_173 (1-123) x US-10-226-956-320 (1-480)
QY 11 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVallysHisPhe 30
Db 190 GTCCGCTCAGGTACAACATGACCCAGGCCACTTCTCCGTTTTATTAGACGTGAACACTTT 249
QY 31 SerProGluAspLeuThrVallysValGlnAspAspPheValGluIleHisGlyLysHis 50
Db 250 AGCCAGAGAGATAGCAGATCAAGTTGTAGGAGAGCATGTGGAAAGTTTCACCGAGACAT 309
QY 51 AsnGluArgGlnAspAspHisGlyTyrlleSerArgGluPheHisArgGlyTyrArgLeu 70

```

Db 310 GAAGAGACACAGATGAACATGTTTCGTAGCGAGAGAAATTCATCGCGGTATCGTCG 369  
QY 71 ProSerAnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 90  
Db 370 CCCCAGAGTCGATCTCGACGTGTGACGAGTCATTCGCTGAGGAGTGCTCAGT 429  
QY 91 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 110  
Db 430 ATC-----CAAGCAGCCCC 444  
QY 111 ValSerArgGluGluLysProThrSerAla 120  
Db 445 CGGTACGCCCAAGCCCGCTTCGGCTGCT 474

## RESULT 46

US-10-226-956-303  
; Sequence 303, Application US/10226956  
; Publication No. US20030060399A1  
; GENERAL INFORMATION:  
; APPLICANT: Brophy, Colleen  
; APPLICANT: Komalavilas, Padmini  
; APPLICANT: Panitch, Alyssa  
; APPLICANT: Joshi, Lokesh  
; APPLICANT: Seal, Brandon L.  
; TITLE OF INVENTION: REAGENTS AND METHODS FOR SMOOTH MUSCLE THERAPIES  
; FILE REFERENCE: ASU-1061-US  
; CURRENT APPLICATION NUMBER: US/10/226,956  
; CURRENT FILING DATE: 2002-08-23  
; PRIOR APPLICATION NUMBER: 60/314,535  
; PRIOR FILING DATE: 2001-08-23  
; NUMBER OF SEQ ID NOS: 320  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 303  
; LENGTH: 486  
; TYPE: DNA  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Sequence encoding rat HSP20

US-10-226-956-303

## Alignment Scores:

Pred. No.:	1,32E-26	Length:	486
Score:	250.50	Matches:	48
Percent Similarity:	63.39%	Conservative:	23
Best Local Similarity:	42.86%	Mismatches:	26
Query Match:	39.20%	Indels:	15
DB:	14	Gaps:	1

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-10-226-956-303 (1-486)

QY 12 SerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPheSer 31  
Db 193 GCCCAAGTCCCGACCGATCCAGGCTATTTCAGCGTCTGTAGACGTAAGCATTTAGT 252  
QY 32 ProGluAspLeuThrValLysValGlnAspPheValGluIleHisGlyLysHisAsn 51  
Db 253 CCAGAGAAATTCAGTAAAGTAGTGGGACCATGTCGAGGTACATGTCAGACGAA 312  
QY 52 GluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeuPro 71  
Db 313 GAGACCTGATGAACACGGTTTCATCGTCGAGAGTTTCACCGCGCTTATCGCTTCCG 372  
QY 72 SerAnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPhe 91  
Db 373 CCGGGGTGATCCCGCGCGCTCATACATCAGTCCGAGGAGGATTTATCCATA 432  
QY 92 CysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProVal 111  
Db 433 -----CAAGCCACACCGGCC 447  
QY 112 SerArgGluGluLysProThrSerAlaProSerSer 123  
Db 448 TCTGCTCAGGCTCGCTTCCATCCCTCCCTCCGCGCA 483

## RESULT 47

US-10-152-319A-1412  
; Sequence 1412, Application US/10152319A  
; Publication No. US20040072160A1  
; GENERAL INFORMATION:  
; APPLICANT: Mendrick, Donna  
; APPLICANT: Porter, Mark  
; APPLICANT: Johnson, Kory  
; APPLICANT: Higgs, Brandon  
; APPLICANT: Castle, Arthur  
; APPLICANT: Elashoff, Michael  
; TITLE OF INVENTION: Molecular Toxicology Modeling  
; FILE REFERENCE: 44921-5089-US  
; CURRENT APPLICATION NUMBER: US/10/152,319A  
; CURRENT FILING DATE: 2002-05-22  
; PRIOR APPLICATION NUMBER: US 60/292,335  
; PRIOR FILING DATE: 2001-05-22  
; PRIOR APPLICATION NUMBER: US 60/297,523  
; PRIOR FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: US 60/298,925  
; PRIOR FILING DATE: 2001-06-19  
; PRIOR APPLICATION NUMBER: US 60/303,810  
; PRIOR FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: US 60/303,807  
; PRIOR FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: US 60/303,808  
; PRIOR FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: US 60/315,047  
; PRIOR FILING DATE: 2001-08-28  
; PRIOR APPLICATION NUMBER: US 60/324,928  
; PRIOR FILING DATE: 2001-09-27  
; PRIOR APPLICATION NUMBER: US 60/330,867  
; PRIOR FILING DATE: 2001-11-01  
; PRIOR APPLICATION NUMBER: US 60/330,462  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 2221  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1412

US-10-152-319A-1412

US-10-152-319A-1412

## Alignment Scores:

Pred. No.:	5.65E-26	Length:	1310
Score:	250.50	Matches:	48
Percent Similarity:	63.39%	Conservative:	23
Best Local Similarity:	42.86%	Mismatches:	26
Query Match:	39.20%	Indels:	15
DB:	17	Gaps:	1

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-10-152-319A-1412 (1-1310)

QY 12 SerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPheSer 31  
Db 198 GCCCAGTCCCGACCGATCCAGGCTATTTCGTGCTGCTGATGTGAAGCATTTCTCG 257  
QY 32 ProGluAspLeuThrValLysValGlnAspPheValGluIleHisGlyLysHisAsn 51  
Db 258 CCAGAGAAATCTCTGTCAAGTGTGTGTCGTCACCATGTGGAGGTCATGCTCGCATGAG 317  
QY 52 GluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeuPro 71  
Db 318 GAGCCCGCAGATGAACATGGATTTCATGTCGAGAGTTCCACCGCATGCTGCTGCT 377  
QY 72 SerAnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPhe 91  
Db 378 CTTGCGGTGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 437

```
Qy 92 CysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProVal 111
Db 438 -----CAGGCCACACCGC 452
Qy 112 SerArgGluGluLysProThrSerAlaProSerSer 123
Db 453 TCGGCCACGGCTCACTTCGCTCACCACCTGCTGCC 488

RESULT 48
US-10-723-860-5075
; Sequence 5075, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE REFERENCE: 05882.0193.NFUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; PRIOR FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5075
; LENGTH: 1820
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-723-860-5075

Alignment Scores:
Pred. No.: 9.16e-26 Length: 1820
Score: 250.50 Matches: 46
Percent Similarity: 63.64% Conservative: 24
Best Local Similarity: 41.82% Mismatches: 25
Query Match: 39.20% Indels: 15
DB: 18 Gaps: 1

US-10-657-740-1_COPY_51_173 (1-123) x US-10-723-860-5075 (1-1820)
Qy 11 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe 30
Db 212 GTCGCCACAGTGGCCGACGCGCCGCCGCTTCGGTGTGCTAGACGTGAAGCACTTC 271
Qy 31 SerProGluAspLeuThrValLysValGlnAspPheValGluIleHisGlyLysHis 50
Db 272 TCGCCGAGGAATGTGTCAAGTGTGTGGCGGACACGTTGGAGGTGCACGCGGCCAC 331
Qy 51 AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgGlyArgLeu 70
Db 332 GAGGAGCGCCCGGATGACACGGATTCTCGCGCGCGAGTTCACCGCTGCTACCGCCTG 391
Qy 71 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 90
Db 392 CCGCTGGCGTGGATTCGGGCTGCGGTGACGTCGCGCTGCCCGCGAGGGGCTGCTGCC 451
Qy 91 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 110
Db 452 ATC-----CAGGCCGCGACCA 466

RESULT 49
US-09-918-995-4949
; Sequence 4949, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: FROM VARIOUS CDNA LIBRARIES
```

```
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4949
; LENGTH: 450
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(450)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-4949

Alignment Scores:
Pred. No.: 1.39e-26 Length: 450
Score: 250.00 Matches: 44
Percent Similarity: 80.52% Conservative: 18
Best Local Similarity: 57.14% Mismatches: 15
Query Match: 39.12% Indels: 0
DB: 10 Gaps: 0

US-10-657-740-1_COPY_51_173 (1-123) x US-09-918-995-4949 (1-450)
Qy 8 AspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 27
Db 220 GACACTGGACTCTCAAGATGCGCTGGAGAGACAGGGTCTCTGTCAACTGGATGTG 279
Qy 28 LysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGluIleHis 47
Db 280 AAGCACTTCTCCCGAGAGAACTCAGAGTTAATGTGTGTGATGATGAGATTGAGGTGCAT 339
Qy 48 GlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArg 67
Db 340 GGAATAACATGAAGAGCGCCAGGATGAACATGGTTTCATCTCCAGGGAGTTCACAGGACA 399
Qy 68 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSer 84
Db 400 TACCGGATCCAGCTGATGATAGACCTCTGACCATTTACTTCACTCCCTGTCN 450

RESULT 50
US-09-911-904-146
; Sequence 146, Application US/09911904
; Publication No. US20030096234A1
; GENERAL INFORMATION:
; APPLICANT: Farr, Spencer B.
; APPLICANT: Pickett, Gavin G.
; APPLICANT: Neft, Robin Eileen
; APPLICANT: Dunn, II, Robert Thomas
; TITLE OF INVENTION: CANINE TOXICITY GENES
; FILE REFERENCE: 40074200200
; CURRENT APPLICATION NUMBER: US/09/911,904
; CURRENT FILING DATE: 2002-04-09
; PRIOR APPLICATION NUMBER: US 60/220,057
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 146
; LENGTH: 503
; TYPE: DNA
; ORGANISM: Canis familiaris
US-09-911-904-146

Alignment Scores:
Pred. No.: 3.23e-26 Length: 503
Score: 248.00 Matches: 53
Percent Similarity: 65.42% Conservative: 17
Best Local Similarity: 49.53% Mismatches: 28
Query Match: 38.81% Indels: 10
DB: 10 Gaps: 1
```

US-10-657-740-1\_COPY\_51\_173 (1-123) x US-09-911-904-146 (1-503)

```
QY 7 LeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAsp 26
Db |||||:::|||||:::| |||||:::|
206 CTCAGCAGCGGCGTCTCGAGATCCGCGACGCGCGACCGCTGGCGCGTCTCCCTGGAC 265
QY 27 ValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGluIle 46
Db |||||:::|||||:::| |||||:::|
266 GTCAACCACTTCGCCCCGAGAGCTGACGGTCAAGACGAGGACGCGCGTGGTGAGATA 325
QY 47 HisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArg 66
Db |||||:::|||||:::| |||||:::|
326 ACTGGCAAGCAGCAAGAGAGGAGGATGAGCATGGTACATCTCCCGCCGCTCACTCCC 385
QY 67 ArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAsp 86
Db ::||| |||||:::| |||||:::|
386 AATACACCCCTGCCCTGGTGTGGATCCTACCTGGTCTCTCTCTCTCTCTCTCTCTCTGAG 445
QY 87 GlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGlu 106
Db ||||| |||||
446 GGCACCTCTCAG-----GTGGAGGCTCC-CATGCCAA 477
QY 107 ArgAlaIleProValSerArg 113
Db ::||| |||||
478 GCCAGCCACCAGTCGGCAGA 498
```

Search completed: May 30, 2005, 09:27:43  
Job time : 873.74 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 30, 2005, 04:00:28 ; Search time 2317.47 Seconds  
(without alignment)

2020.266 Million cell updates/sec

Title: US-10-657-740-1\_COPY\_51\_173

Perfect score: 639

Sequence: 1 SLFRVLDGSEVRSRDK.....HAERAIQVRSREKETSAPSS 123

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO\_spool/US10657740/runat\_27052005\_165253\_3329/app\_query.fasta\_1.590  
-DB=EST -OPMT=fastcap -SUFFIX=first -MINMATCH=0.1 -IOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=150  
-DOCALIGN=200 -THRM score=ptc -THR MAX=100 -THR MIN=0 -ALIGN=50 -MODE=LOCAL  
-OUTPMT=ptc -NORM=ext -HRAPISE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10657740 @CGN 1 1 5533 @runat\_27052005\_165253\_3329 -NCPUP=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORE=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAP=10 -XGAPEXT=0.5 -FCGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:  
1: gb\_est1:  
2: gb\_est2:  
3: gb\_hc:  
4: gb\_est3:  
5: gb\_est4:  
6: gb\_est5:  
7: gb\_est6:  
8: gb\_ga1:  
9: gb\_ga2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	639	100.0	438	2	BF727355
2	639	100.0	446	4	BM697673
3	639	100.0	460	2	BF727464
4	639	100.0	466	2	BF727220
5	639	100.0	500	2	BF727239
6	639	100.0	508	4	BM696489
7	639	100.0	511	4	BM697090
8	639	100.0	512	4	BM697099
9	639	100.0	518	2	BF726854

10	639	100.0	522	9	AY419529	Homo sapi
11	639	100.0	523	2	BF726890	by13f01.y
12	639	100.0	533	4	BM704974	UI-E-DWO-
13	639	100.0	534	4	BM705774	UI-E-DWO-
14	639	100.0	536	2	BF726399	by06a09.y
15	639	100.0	558	2	BF726679	by10d02.y
16	639	100.0	569	4	BM697368	UI-E-DWO-
17	639	100.0	576	2	BF726236	by03f01.y
18	639	100.0	577	2	BF727002	by15c11.y
19	639	100.0	578	4	BM721893	UI-E-E00-
20	639	100.0	592	2	BF727295	by19e10.y
21	639	100.0	592	4	BM705926	UI-E-DWO-
22	639	100.0	593	4	BM697101	UI-E-DWO-
23	639	100.0	596	4	BM696581	UI-E-DWO-
24	639	100.0	599	4	BM696976	UI-E-DWO-
25	639	100.0	607	6	CD675250	fs21c02.y
26	639	100.0	612	6	CD676130	fs32e02.y
27	639	100.0	629	2	BF727028	by15g05.y
28	639	100.0	630	4	BM696651	UI-E-DWO-
29	639	100.0	631	4	BM706251	UI-E-DWO-
30	639	100.0	633	6	CD672144	fg10b07.y
31	639	100.0	635	4	BM697066	UI-E-DWO-
32	639	100.0	659	4	BM706270	UI-E-DWO-
33	639	100.0	661	4	BM697160	UI-E-DWO-
34	639	100.0	681	5	EX118596	EX118596
35	639	100.0	698	2	BF727324	by19h12.y
36	639	100.0	724	4	BM722336	UI-E-E00-
37	635	99.4	559	4	BM722779	UI-E-E00-
38	633	99.1	532	2	BF726358	by05d12.y
39	632	98.9	522	9	AY419530	Pan trogl
40	631	98.7	519	4	BM696477	UI-E-DWO-
41	630	98.6	577	2	BF726253	by03h05.y
42	630	98.6	587	2	BF726422	by06d05.y
43	629	98.4	597	2	BF726330	by05b01.y
44	626	98.4	604	5	BQ640267	he26c10.y
45	626	98.0	580	4	BM706139	UI-E-DWO-
46	626	98.0	586	4	BM722650	UI-E-E00-
47	626	98.0	607	4	BM696670	UI-E-DWO-
48	620	97.0	512	4	BM706160	UI-E-DWO-
49	620	97.0	544	2	BF726363	by05e10.y
50	616	96.4	498	4	BM696853	UI-E-DWO-
51	616	96.4	523	2	BF726438	by06f01.y
52	616	96.4	588	4	BM696799	UI-E-DWO-
53	615	96.2	663	4	BM686748	UI-E-CQ0-
54	611	95.6	371	4	BM705517	UI-E-DX0-
55	611	95.6	503	2	BF727205	by18b11.y
56	611	95.6	558	4	BM696667	UI-E-DWO-
57	601	94.1	688	7	CF732528	UI-M-HA0-
58	600	93.9	536	4	BM686206	UI-E-C11-
59	599	93.7	447	6	CB848427	M2PN-4253
60	599	93.7	486	7	CK627569	io05f11.y
61	599	93.7	522	9	AY419531	Mus muscu
62	599	93.7	542	5	BQ637175	he06f05.y
63	599	93.7	552	7	CK628252	ip09h08.y
64	599	93.7	556	6	CB848450	M2PN-4276
65	599	93.7	557	7	CK627503	io03g04.y
66	599	93.7	558	7	CK627631	io06f03.y
67	599	93.7	566	7	CK627600	io06c02.y
68	599	93.7	570	7	CK628219	ip09d07.y
69	599	93.7	574	7	CK627249	io01b06.y
70	599	93.7	579	7	CK628200	ip09b11.y
71	599	93.7	581	7	CO429470	UI-N-HW0-
72	599	93.7	582	7	CK628246	ip09g11.y
73	599	93.7	582	7	CK628391	ip11f11.y
74	599	93.7	586	7	CK627945	ip04h08.y
75	599	93.7	588	7	CK628525	ip13f04.y
76	599	93.7	590	7	CK627884	ip04a08.y
77	599	93.7	590	7	CK628336	ip10h08.y
78	599	93.7	590	7	CK628526	ip13f05.y
79	599	93.7	593	7	CK628271	ip10b05.y
80	599	93.7	594	7	CK628372	ip11d12.y
81	599	93.7	600	4	BG805475	0693-23 M
82	599	93.7	600	4	BG808732	2121-59 M

```

83 599 93.7 600 4 BI991006
84 599 93.7 600 6 CA527708
85 599 93.7 601 7 CK627483
86 599 93.7 601 7 CK627500
87 599 93.7 601 7 CK627510
88 599 93.7 602 7 CK627644
89 599 93.7 606 6 CB849766
90 599 93.7 606 6 CK628546
91 599 93.7 615 7 CK627740
92 599 93.7 617 6 CB848065
93 599 93.7 621 6 CB845999
94 599 93.7 622 7 CK627736
95 599 93.7 623 7 CK627777
96 599 93.7 626 7 CK627695
97 599 93.7 626 7 CK627737
98 599 93.7 629 6 CB055541
99 599 93.7 632 6 CB842313
100 599 93.7 633 7 CK627814
101 599 93.7 636 7 CK627616
102 599 93.7 636 7 CK627776
103 599 93.7 637 6 CB848191
104 599 93.7 639 7 CK627374
105 599 93.7 642 7 CK627318
106 599 93.7 643 6 CB846376
107 599 93.7 643 7 CK627445
108 599 93.7 643 7 CK629892
109 599 93.7 645 6 CB848071
110 599 93.7 645 7 CK627651
111 599 93.7 659 6 CB841917
112 599 93.7 662 6 CB841336
113 599 93.7 665 6 CB849833
114 599 93.7 669 6 CB840463
115 599 93.7 669 6 CB840876
116 599 93.7 676 6 CB840784
117 599 93.7 688 7 CO427248
118 599 93.7 690 6 CB846475
119 599 93.7 690 7 CN455827
120 599 93.7 690 7 CO428179
121 599 93.7 691 7 CF731811
122 599 93.7 692 6 CB845729
123 599 93.7 696 7 CO429233
124 599 93.7 699 7 CO427006
125 599 93.7 699 7 CO427334
126 599 93.7 702 6 CB844707
127 599 93.7 705 7 CO429072
128 599 93.7 705 7 CO429948
129 599 93.7 709 7 CO428729
130 599 93.7 721 7 CF733985
131 599 93.7 721 7 CF731519
132 599 93.7 760 7 CO427314
133 599 93.7 763 7 CO427704
134 599 93.7 771 7 CO427433
135 599 93.7 773 7 CO428121
136 599 93.7 798 7 CO427114
137 599 93.7 834 6 CB845009
138 599 93.7 851 7 CB845124
139 599 93.6 677 7 CN435619
140 598 93.6 745 7 CN441369
141 598 93.6 809 7 CN438019
142 598 93.6 836 7 CR550416
143 598 93.6 897 7 CR550671
144 597 93.4 618 4 BM697176
145 595 93.1 573 4 CB844502
146 595 93.1 698 7 CO429320
147 594 93.0 478 4 BM705752
148 594 93.0 854 6 CB848776
149 593 92.8 619 7 CK627238
150 593 92.8 625 7 CK627598

```

## ALIGNMENTS

```

BF727355
LOCUS
DEFINITION
  by20d03.y1 Human Lens cDNA (Un-normalized, unamplified): BY Homo
  sapiens cDNA clone by20d03 5', mRNA sequence.
ACCESSION
  BF727355
VERSION
  BF727355.1 GI:12043266
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 438)
REFERENCE
  Wisotzky G., Bernstein S., Behal A. and Smith D.
  NEIBANK: EST analysis and bioinformatics for ocular genomics
  Invest. Ophthalmol. Vis. Sci. 41 (2000) In press
  Contact: Wisotzky G
  Section on Molecular Structure and Function
  National Eye Institute
  6/331, NIH, Bethesda, MD 20892-2740, USA
  Tel: 301 402 3452
  Fax: 301 496 0078
  Email: graeme@helix.nih.gov
  Plate: 20 row: d column: 03
  Seq primer: M13Rpl reverse primer (ABI).
  Location/Qualifiers
    1..438
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="by20d03"
      /tissue_type="Lens"
      /dev_stage="Adult"
      /lab_host="EMDH10B"
      /clone_lib="Human Lens cDNA (Un-normalized, unamplified):
      BY"
      /notes="Organ: Eye; Vector: pCMVSPORT6; Two human lenses
      from different adults (both approximately 40 years old)
      together yielded 20ug of total RNA and 150ng mRNA for cDNA
      library synthesis. A directionally cloned cDNA library in
      the pCMVSPORT6 vector was constructed at Life
      Technologies, essentially following the protocols of the
      SuperScript Plasmid System full details of which are
      contained in the manufacturer's instruction manual
      (http://www.lifetech.com/). First strand synthesis was
      carried out using a Not I primer-adaptor
      [5'-pGACAGTCTAGATCGGAGCGGCCCT(T)15-3']. Not I/blunt
      end inserts were cloned into the Not I/EcoR V sites in the
      vector. EST analysis was performed on the unamplified
      library at the NIH Intramural Sequencing Center (NISC)."
ORIGIN
Alignment Scores:
Pred. No.: 3.16e-68 Length: 438
Score: 639.00 Matches: 123
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 2 Gaps: 0
US-10-657-740-1_COPY_51_173 (1-123) x BF727355 (1-438)
QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20
Db 7 TCCTCTTCGACCGCTGCTGGACTCGGACCTCTCTGAGTTTCGATCCGACCGACAG 66
QY 21 PheValIlePheLeuAspValIshisPheSerProGluAspLeuThrValGln 40
Db 67 TTCGTCATCTTCCTCGATGTGAAGCACTTCTCCCGAGGACCTCACCGTGAAGTGCAG 126
QY 41 AspAspPheValGluIleHisGlyIshisGlyIshisGluArgGlnAspAspHisGlyTyrile 60
Db 127 GACGACTTTGTGGATCCACGGAAAGCACACGAGCGCCGACGACGACGACGACGACGAC 186

```



QY 61 SerArgGluPheHisArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
 Db 187 TCCCGGAGTTCCACCGCGCTACCGCGCTCCACGTCGACAGTCGCGCCCTCTCT 246

QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100  
 Db 247 TGCTCCCTGTCTGCCGATGGCATGCTGACCTTCTGTGGCCCCCAAGATCCAGACTGGCCTG 306

QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120  
 Db 307 GATGCCACCCAGCGGAGCGCATCCCGGTGTGCGGGAGGAGAGCCACCTCGGCT 366

QY 121 ProSerSer 123  
 Db 367 CCTCGTCC 375

RESULT 2  
 BM697673  
 LOCUS  
 DEFINITION  
 UI-E-DX0-agn-d-18-0-UI.r1 UI-E-DX0 Homo sapiens cDNA clone  
 UI-E-DX0-agn-d-18-0-UI 5', mRNA sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 446)  
 Ronaldo,M.F., Lennon,G. and Soares,M.B.  
 Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 Genome Res. 6 (9), 791-806 (1996)  
 97044477  
 8889548  
 Contact: Soares, MB  
 Coordinated Laboratory for Computational Genomics  
 University of Iowa  
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: bento-soares@uiowa.edu  
 Tissue Procurement: Dr. Gregg Hageman  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research  
 Genetics (www.resgen.com).  
 Seq primer: M13 Reverse.

FEATURES  
 source  
 1..446  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-E-DX0-agn-d-18-0-UI"  
 /tissue\_type="fetal eyes"  
 /dev\_stage="fetal"  
 /lab\_host="pH10B (Life Technologies) (T1 phage resistant)"  
 /clone\_lib="UI-E-DX0"  
 /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a  
 modified polylinker; Site 1: EcoR I; Site 2: Not I;  
 UI-E-DX0 is a cDNA library containing the following  
 tissue(s): fetal eyes. The library was constructed  
 according to Ronaldo, Lennon and Soares, Genome Research,  
 6:791-806, 1996. First strand cDNA synthesis was primed  
 with an oligo-dT primer containing a Not I site. Double  
 stranded cDNA was ligated to an EcoR I adaptor, digested  
 with Not I, and cloned directionally into pT7T3-Pac  
 vector. The oligonucleotide used to prime the synthesis of  
 first-strand cDNA contains a library tag sequence that is  
 located between the Not I site and the (dT)18 tail. The  
 sequence tag for this library is AGAATCAAGA. This library  
 was created for the program, Gene Discovery in the Visual

System, supported by National Eye Institute (NEI)."

ORIGIN  
 Alignment Scores:  
 Pred. No.: 3.23e-68 Length: 446  
 Score: 639.00 Matches: 123  
 Percent Similarity: 100.00% Conservatives: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 4 Gaps: 0  
 US-10-657-740-1\_COPY\_51\_173 (1-123) x BM697673 (1-446)

QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20  
 Db 56 TCCCTCTTCCGACCGCTGCTGACTCCGACATCTCTGAGGTTTCGATCCGACCGGCAAG 115

QY 21 PheValIlePheLeuAspVallyHisPheSerProGluAspLeuThrVallyValGln 40  
 Db 116 TTGCGATCTTCTCTCGATGTGAAGCACCTTCTCCCGAGGAGCTCACCGTGAAGGTGAG 175

QY 41 AspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspHisGlyTyrIle 60  
 Db 176 GACGACTTTGTGGAGATCCACGGAAGACACACGAGCGCCAGGACGACCGCTACATT 235

QY 61 SerArgGluPheHisArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
 Db 236 TCCCGTGAGTTCCACCGCGCTACCGCTCGCTCCACGTCGAGCCAGTCCGCTCTCT 295

QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100  
 Db 296 TGCTCCCTGTCTGCCGATGGCATGCTGACCTTCTGTGGCCCCCAAGATCCAGACTGGCCTG 355

QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120  
 Db 356 GATGCCACCCAGCGGAGCGCATCCCGGTGTGCGGGAGGAGAGGCCACCTCGGCT 415

QY 121 ProSerSer 123  
 Db 416 CCTCGTCC 424

RESULT 3  
 BM697673  
 LOCUS  
 DEFINITION  
 by21h03.y1 Human Lens cDNA (Un-normalized, unamplified): BY Homo  
 sapiens cDNA clone by21h03 5', mRNA sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 460)  
 Wistow,G.J., Bernstein,S., Behal,A. and Smith,D.  
 NEIBANK: EST analysis and bioinformatics for ocular genomics  
 Invest. Ophthalmol. Vis. Sci. 41 (2000) In press  
 Contact: Wistow G  
 Section on Molecular Structure and Function  
 National Eye Institute  
 6/331, NIH, Bethesda, MD 20892-2740, USA  
 Tel: 301 402 3452  
 Fax: 301 496 0078  
 Email: graeme@helix.nih.gov  
 Plate: 21 row: h column: 03  
 Seq primer: M13R1 reverse primer (ABI).  
 Location/Qualifiers  
 1..460  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="by21h03"  
 /tissue\_type="Lens"

/dev\_stage="Adult"  
/lab\_host="EMDH10B"  
By  
/note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses from different adults (both approximately 40 years old) together yielded 20ug of total RNA and 150ng mRNA for cDNA library synthesis. A directionally cloned cDNA library in the pCMVSPORT6 vector was constructed at Life Technologies, essentially following the protocols of the Superscript Plasmid System full details of which are contained in the manufacturer's instruction manual (http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adaptor [5'-pGACTAGTTCTAGATCGGCGGCCCT(T)15-3']. Not I/blunt end inserts were cloned into the Not I/EcoR V sites in the vector. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

/clone\_lib="Human Lens cDNA (Un-normalized, unamplified):

/note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses from different adults (both approximately 40 years old) together yielded 20ug of total RNA and 150ng mRNA for cDNA library synthesis. A directionally cloned cDNA library in the pCMVSPORT6 vector was constructed at Life Technologies, essentially following the protocols of the Superscript Plasmid System full details of which are contained in the manufacturer's instruction manual (http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adaptor [5'-pGACTAGTTCTAGATCGGCGGCCCT(T)15-3']. Not I/blunt end inserts were cloned into the Not I/EcoR V sites in the vector. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

/dev\_stage="Adult"  
/lab\_host="EMDH10B"  
By  
/note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses from different adults (both approximately 40 years old) together yielded 20ug of total RNA and 150ng mRNA for cDNA library synthesis. A directionally cloned cDNA library in the pCMVSPORT6 vector was constructed at Life Technologies, essentially following the protocols of the Superscript Plasmid System full details of which are contained in the manufacturer's instruction manual (http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adaptor [5'-pGACTAGTTCTAGATCGGCGGCCCT(T)15-3']. Not I/blunt end inserts were cloned into the Not I/EcoR V sites in the vector. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

/dev\_stage="Adult"  
/lab\_host="EMDH10B"  
By  
/note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses from different adults (both approximately 40 years old) together yielded 20ug of total RNA and 150ng mRNA for cDNA library synthesis. A directionally cloned cDNA library in the pCMVSPORT6 vector was constructed at Life Technologies, essentially following the protocols of the Superscript Plasmid System full details of which are contained in the manufacturer's instruction manual (http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adaptor [5'-pGACTAGTTCTAGATCGGCGGCCCT(T)15-3']. Not I/blunt end inserts were cloned into the Not I/EcoR V sites in the vector. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

## ORIGIN

Alignment Scores:  
Pred. No.: 3,37e-68 Length: 460  
Score: 639.00 Matches: 123  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 2 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x BF727464 (1-460)

QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerArgAspLys 20  
Db 21 TCCCTCTTCGCGACCGTGGATCTCGGCATCTCTGAGTTCGATCCGCGGACAG 80  
QY 21 PheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrValLysValGln 40  
Db 81 TTGCTCATCTTCTCGATGTGAAGCACTTCTCCCGAGAGGACCTCACCGTGAAGGTGCAG 140  
QY 41 AspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIle 60  
Db 141 GAGGACTTTGTGGAGATCCAGCGAAGACACACGAGCGCGGACGACGACCGGCTACATT 200  
QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
Db 201 TCCCGTGAAGTTCACCGCGCGCTACCGCGCTCGCGTCCGCGTCAACGTGAGGACGCTCTCT 260  
QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100  
Db 261 TGCTCCCTGTCTGCGGATGGATGCTGACCTTCTGTGCGCCCAAGATCGAGCTGGCTG 320  
QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120  
Db 321 GATGCCACCCAGCGGAGGAGCCATCCCGTGTGCGGAGGAGGAGGAGGAGGAGGAGGAG 380  
QY 121 ProSerSer 123  
Db 381 CCCTGCTCC 389

## RESULT 4

BF727220

LOCUS

DEFINITION

BF727220

ACCESION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 466)

Wistow G.J., Bernstein, S., Behal, A. and Smith, D.

NEIBANK: EST analysis and bioinformatics for ocular genomics

## JOURNAL

COMMENT

Invest. Ophthalmol. Vis. Sci. 41 (2000) In press  
Contact: Wistow G  
Section on Molecular Structure and Function  
National Eye Institute  
Tel: 301 402 3452  
Fax: 301 496 0078  
Email: graeme@helix.nih.gov  
Plate: 18 row: d column: 12  
Seq primer: M13RF1 reverse primer (ABI).  
Location/Qualifiers  
1. .466  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="by18d12"  
/tissue\_type="Lens"  
/dev\_stage="Adult"  
/lab\_host="EMDH10B"  
/clone\_lib="Human Lens cDNA (Un-normalized, unamplified):  
By"

## FEATURES

source

1. .466

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="by18d12"

/tissue\_type="Lens"

/dev\_stage="Adult"

/lab\_host="EMDH10B"

/clone\_lib="Human Lens cDNA (Un-normalized, unamplified):  
By"

/note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses from different adults (both approximately 40 years old) together yielded 20ug of total RNA and 150ng mRNA for cDNA library synthesis. A directionally cloned cDNA library in the pCMVSPORT6 vector was constructed at Life Technologies, essentially following the protocols of the Superscript Plasmid System full details of which are contained in the manufacturer's instruction manual (http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adaptor [5'-pGACTAGTTCTAGATCGGCGGCCCT(T)15-3']. Not I/blunt end inserts were cloned into the Not I/EcoR V sites in the vector. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

## ORIGIN

Alignment Scores:  
Pred. No.: 3,42e-68 Length: 466  
Score: 639.00 Matches: 123  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 2 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x BF727220 (1-466)

QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerArgAspLys 20  
Db 29 TCCCTCTTCGCGACCGTGGATCTCGGCATCTCTGAGTTCGATCCGCGGACAG 88  
QY 21 PheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrValLysValGln 40  
Db 89 TTGCTCATCTTCTCGATGTGAAGCACTTCTCCCGAGGACCTCACCGTGAAGGTGCAG 148  
QY 41 AspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIle 60  
Db 149 GAGGACTTTGTGGAGATCCAGCGAAGACACACGAGCGCGGACGACGACCGGCTACATT 208  
QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
Db 209 TCCCGTGAAGTTCACCGCGCTTACCGCTGCGTCCACGCGGACGAGCTCGGCCCTCTCT 268  
QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100  
Db 269 TGCTCCCTGTCTGCGGATGGATGCTGACCTTCTGTGCGCCCAAGATCCAGACTGGCTG 328  
QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120  
Db 329 GATGCCACCCAGCGGAGGAGCCATCCCGTGTGCGGAGGAGGAGGAGGAGGAGGAGGAG 388  
QY 121 ProSerSer 123

```

Db          389 CCCTCGTCC 397

RESULT 5
BF727239
LOCUS
DEFINITION
  BF727239 500 bp mRNA linear EST 05-JAN-2001
  sapiens cDNA clone by18g02 5', mRNA sequence.
ACCESSION
BF727239
VERSION
BF727239.1 GI:12043150
KEYWORDS
EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 500)
REFERENCE
  Wistow,G.J., Bernstein,S., Behal,A. and Smith,D.
  NEIBANK: EST analysis and bioinformatics for ocular genomics
  Invest. Ophthalmol. Vis. Sci. 41 (2000) In press
  Contact: Wistow G
COMMENT
  Section on Molecular Structure and Function
  National Eye Institute
  6/331, NIH, Bethesda, MD 20892-2740, USA
  Tel: 301 402 3452
  Fax: 301 496 0078
  Email: graeme@helix.nih.gov
  Plate: 18 row: g column: 02
  Seq primer: M13RPI reverse primer (ABI).
  Location/Qualifiers
    1..500
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="by18g02"
      /tissue type="Lens"
      /dev stage="Adult"
      /lab_host="EMDH108"
      /clone_lib="Human Lens cDNA (Un-normalized, unamplified):
      By"
      /note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses
      from different adults (both approximately 40 years old)
      together yielded 20ug of total RNA and 150ng mRNA for cDNA
      library synthesis. A directionally cloned cDNA library in
      the pCMVSPORT6 vector was constructed at Life
      Technologies, essentially following the protocols of the
      SuperScript Plasmid System full details of which are
      contained in the manufacturer's instruction manual
      (http://www.lifetech.com/). First strand synthesis was
      carried out using a Not I primer-adaptor
      [5'-pGACTAGTTTATGTCGAGCGGCCGCC(T)15-3']. Not I/blunt
      end inserts were cloned into the Not I/EcoR V sites in the
      vector. EST analysis was performed on the unamplified
      library at the NIH Intramural Sequencing Center (NISC)."
```

ORIGIN

Alignment Scores:

Pred. No.:	3.75e-68	Length:	500
Score:	639.00	Matches:	123
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
Db:	2	Gaps:	0

US-10-657-740-1\_COPY\_51\_173 (1-123) x BF727239 (1-500)

Qy 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspAspIys 20

Db 7 TCCTCTTCCGACCGCTCGATCCGCGCATCTTGAGGTTCGATCCGACCGGACAG 66

Qy 21 PheValIlePheLeuAspValIysHisPheSerProGluAspLeuThrValIysValGln 40

Db 67 TTGTCATCTCTCTCGATCGAGCATTCTCCCGAGGAGGACCTCACCCTGAGGTGCAG 126

Qy 41 AspAspPheValGluIleHisGlyIysHisAsnGluArgGlnAspAspHisGlyTyrIle 60

Db 127 GACGACTTTGTGGAGATCCAGGAAGCACACGAGCGCCAGCACCGCTACAT 186

Qy 61 SerArgGluPheHisArgArgTyArgLeuProSerAsnValAspGlnSerAlaLeuSer 80

Db 187 TCCCGTGAGTTCACCGCCGCTACCGCTCGGTCACAGGTGGACAGTCGGCCCTCT 246

Qy 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100

Db 247 TGCTCCCTGTCTGCCGATGGCATGTGACCTTCTGTGGCCCCCAGATCCAGATCGCCTG 306

Qy 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120

Db 307 GATGCCACCCAGCGGAGCGGATCCCTGCTGCGGGAGAGAGAGCCACCTCGGCT 366

Qy 121 ProSerSer 123

Db 367 CCCTCGTCC 375

RESULT 6

BM696489

LOCUS

DEFINITION

UI-E-DW0-agj-n-20-0-UI.r1 UI-E-DW0 Homo sapiens cDNA clone

UI-E-DW0-agj-n-20-0-UI 5', mRNA sequence.

ACCESSION

BM696489

VERSION

BM696489.1 GI:19009747

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 508)

REFERENCE

AUTHORS

Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE

Normalization and subtraction: two approaches to facilitate gene

discovery

JOURNAL

Genome Res. 6 (9), 791-806 (1996)

MEDLINE

97044477

PUBMED

8889548

COMMENT

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

Tissue Procurement: Dr. Gregg Hageman

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research

Genetics (www.resgen.com).

Seq primer: M13 Reverse.

Location/Qualifiers

1..508

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="UI-E-DW0-agj-n-20-0-UI"

/tissue type="lens"

/dev stage="adult"

/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"

/clone\_lib="UI-E-DW0"

/note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a

modified polylinker; Site 1: EcoR I; Site 2: Not I;

UI-E-DW0 is a cDNA library containing the following

tissue(s): lens. The library was constructed according to

Bonaldo, Lennon and Soares, Genome Research, 6:791-806,

1996. First strand cDNA synthesis was primed with an

oligo-dT primer containing a Not I site. Double stranded

cDNA was ligated to an EcoR I adaptor, digested with Not

I, and cloned directionally into pT7T3-Pac vector. The

oligonucleotide used to prime the synthesis of

first-strand cDNA contains a library tag sequence that is

located between the Not I site and the (dT)18 tail. The sequence tag for this library is CGATTAGCGA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

## ORIGIN

## Alignment Scores:

Pred. No.: 3,83e-68 Length: 508  
Score: 639.00 Matches: 123  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 4 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x BM696489 (1-508)

QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValAla: gSerAspArgAspLys 20  
Db 133 TCCCTCTTCGACCGTGTGGACTCCGGCATCTCTGAGGTTGATCCGACCGGACAAG 192  
QY 21 PheValIlePheLeuAspValIleHisPheSerProGluAspLeuThrValIleValGln 40  
Db 193 TTGCTCATCTTCTCGATGTGAAGCACTTCTCCCGGAGGACCTCAACGGAAGGTGCAG 252  
QY 41 AspAspPheValGluIleHisGlyIleHisAsnGluArgGlnAspAspHisGlyIle 60  
Db 253 GACGACTTTGTGGAGATCCAGGAAGCAACAGCGCCAGGACGACCACTACATT 312  
QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
Db 313 TCCCGTGAGTTCCACCGCGCTACCGCTGCCGTCCACGTGGACCACTGCGGCCCTCTCT 372  
QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100  
Db 373 TGCTCCCTGTGTCGATGGCATGCTGACCTTCTGTGGCCCAAGATCCAGACTGGCCTG 432  
QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120  
Db 433 GATGCCACCCAGCGCGAGCGATCCCGTGTCCGGGAGAGAGAGAGAGAGAGAGAGAG 492  
QY 121 ProSerSer 123  
Db 493 CCCTCGTCC 501

## RESULT 7

BM697090

## LOCUS

DEFINITION UI-E-DW0-0-UI.r1 UI-E-DW0 Homo sapiens linear EST 28-FEB-2002  
UI-E-DW0-0-UI.r1 UI-E-DW0 Homo sapiens cdna clone

## ACCESSION

BM697090

## VERSION

EST.

## KEYWORDS

BM697090.1 GI:19010348

## SOURCE

Homo sapiens

## ORGANISM

Homo sapiens (human)

## REFERENCE

1 (bases 1 to 511)

## AUTHORS

Bonaldo, M.F., Lennon, G. and Soares, M.B.

## TITLE

Normalization and subtraction: two approaches to facilitate gene

## JOURNAL

discovery

## MEDLINE

Genome Res. 6 (9), 791-806 (1996)

## PUBMED

97044477

## COMMENT

8889548

## CONTACT

Contact: Soares, MB

## COORDINATED

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

Tissue Procurement: Dr. Gregg Hageman

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics (www.regen.com).  
Seq primer: M13 Reverse.

## FEATURES

## source

1..511

Location/Qualifiers

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="UI-E-DW0-0-UI"

/tissue\_type="lens"

/dev\_stage="adult"

/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"

/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a

modified polylinker; Site 1: EcoR I; Site 2: Not I;

UI-E-DW0 is a cDNA library containing the following

tissue(s): lens. The library was constructed according to

Bonaldo, Lennon and Soares, Genome Research, 6:791-806,

1996. First strand cDNA synthesis was primed with an

oligo-dT primer containing a Not I site. Double stranded

cDNA was ligated to an EcoR I adaptor, digested with Not

I, and cloned directionally into pT73-Pac vector. The

oligonucleotide used to prime the synthesis of

first-strand cDNA contains a library tag sequence that is

located between the Not I site and the (dT)18 tail. The

sequence tag for this library is CGATTAGCGA. This library

was created for the program, Gene Discovery in the Visual

System, supported by National Eye Institute (NEI)."

ORIGIN

Alignment Scores:

Pred. No.: 3,86e-68 Length: 511

Score: 639.00 Matches: 123

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 4 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x BM697090 (1-511)

QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValAla: gSerAspArgAspLys 20

Db 18 TCCCTCTTCGACCGTGTGGACTCCGGCATCTCTGAGGTTGATCCGACCGGACAAG 77

QY 21 PheValIlePheLeuAspValIleHisPheSerProGluAspLeuThrValIleValGln 40

Db .78 TTGCTCATCTTCTCGATGTGAAGCACTTCTCCCGGAGGACCTCACCGTGAAGTGCAG 137

QY 41 AspAspPheValGluIleHisGlyIleHisAsnGluArgGlnAspAspHisGlyIle 60

Db 138 GACGACTTTGTGGAGATCCAGGAAGCAACAGCGCCAGGACGACCACTACATT 197

QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80

Db 198 TCCCGTGAGTTCCACCGCGCTACCGCTGCCGTCCACCGTGGACCACTGCGGCCCTCTCT 257

QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100

Db 258 TGCTCCCTGTGTCGATGGCATGCTGACCTTCTGTGGCCCAAGATCCAGACTGGCCTG 317

QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120

Db 318 GATGCCACCCAGCGCGAGCGACCATCCCGTGTCCGGGAGAGAGAGAGAGAGAGAGAG 377

QY 121 ProSerSer 123

Db 378 CCCTCGTCC 386

RESULT 8

BM697099

## LOCUS

DEFINITION UI-E-DW0-0-UI.r1 UI-E-DW0 Homo sapiens linear EST 28-FEB-2002



```

Alignment Scores:
Pred. No.: 3,93e-68 Length: 518
Score: 639.00 Matches: 123
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 2 Gaps: 0

US-10-657-740-1_COPY_51_173 (1-123) x BF726854 (1-518)
QY 1 SerLeuPheAtrgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20
Db 140 TCCCTCTCCGACCGTCTGGACTCCGGCATCTCTGAGTTCCATCCGACCGGACAAG 199
QY 21 PheValIlePheLeuAspValIleHisPheSerProGluAspLeuThrValIleValGln 40
Db 200 TTGTCATCTTCTCGATGTGAAGCATCTTCTCCCGGAGGACCTTCACCGTGAAGGTGCAG 259
QY 41 AspAspPheValGluIleHisGlyIleHisAsnGluArgGlnAspAspHisGlyIle 60
Db 260 GACGACTTTGTGGAGATCCACGGAAGCACACGAGCGCCAGGACGACGCTACATT 319
QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80
Db 320 TCCCGTGAGTTCACCCCGCTACCGCTCCGCTCCAACTGACGACGCTCGGCTCTCT 379
QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100
Db 380 TGCTCCCTGTCTGCGATGTCATGCTGACCTTCTGTGGCCCCAAGATCCAGACTGCGCTG 439
QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluLulysProThrSerAla 120
Db 440 GATGCCACCCACCGCGAGGAGCATCCCGTGTGCGGAGGAGAGGCCACCTCGGCT 499
QY 121 ProSerSer 123
Db 500 CCCTCGTCC 508

RESULT 10
AY419529 522 bp DNA linear GSS 17-DEC-2003
LOCUS Homo sapiens CRYAA gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY419529
VERSION AY419529.1 GI:39775486
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
AUTHORS Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
COMMENT Rockville, MD 20850, USA
These sequences were made by sequencing genomic exons and ordering
them based on alignment.
FEATURES Location/Qualifiers
source 1..522
/organism="Homo sapiens"
/mol_type="genomic DNA"

/db_xref="taxon:9606"
<1..>522
/genes="CRYAA"
/locus_tag="HCM6921"

ORIGIN
Alignment Scores:
Pred. No.: 3,97e-68 Length: 522
Score: 639.00 Matches: 123
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-10-657-740-1_COPY_51_173 (1-123) x AY419529 (1-522)
QY 1 SerLeuPheAtrgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20
Db 151 TCCCTCTTCCGACCGTCTGGACTCCGGCATCTCTGAGTTCCATCCGACCGGACAAG 210
QY 21 PheValIlePheLeuAspValIleHisPheSerProGluAspLeuThrValIleValGln 40
Db 211 TTGTCATCTTCTCGATGTGAAGCATCTTCTCCCGGAGGACCTTCACCGTGAAGGTGCAG 270
QY 41 AspAspPheValGluIleHisGlyIleHisAsnGluArgGlnAspAspHisGlyIle 60
Db 271 GACGACTTTGTGGAGATCCACGGAAGCACACGAGCGCCAGGACGACGCTACATT 330
QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80
Db 331 TCCCGTGAGTTCACCCCGCTACCGCTCCGCTCCAACTGACGACGCTCGGCTCTCT 390
QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100
Db 391 TGCTCCCTGTCTGCGATGTCATGCTGACCTTCTGTGGCCCCAAGATCCAGACTGCGCTG 450
QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluLulysProThrSerAla 120
Db 451 GATGCCACCCACCGCGAGGAGCATCCCGTGTGCGGAGGAGAGGCCACCTCGGCT 510
QY 121 ProSerSer 123
Db 511 CCCTCGTCC 519

RESULT 11
BF726890 523 bp mRNA linear EST 05-JAN-2001
LOCUS sapiens cDNA (Un-normalized, unamplified): BY Homo
DEFINITION sapiens cDNA clone by13f01 5', mRNA sequence.
ACCESSION BF726890
VERSION BF726890.1 GI:12042801
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Wistow,G.J., Bernstein,S., Behal,A. and Smith,D.
TITLE NEIBANK: EST analysis and bioinformatics for ocular genomics
JOURNAL Invest. Ophthalmol. Vis. Sci. 41 (2000) In press
COMMENT Contact: Wistow G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graeme@helix.nih.gov
Place: 13 row: f column: 01
Seq primer: M13RP1 reverse primer (ABI).
FEATURES Location/Qualifiers
source 1..523
/organism="Homo sapiens"
/mol_type="mRNA"

```

/db\_xref="taxon:9606"  
 /clone="by33f01"  
 /tissue\_type="Lens"  
 /dev\_stage="Adult"  
 /lab\_host="EMDH103"  
 /clone\_lib="Human Lens cDNA (Un-normalized, unamplified):  
 BY"  
 /note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses from different adults (both approximately 40 years old) together yielded 20ug of total RNA and 150ng mRNA for cDNA library synthesis. A directionally cloned cDNA library in the pCMVSPORT6 vector was constructed at Life Technologies, essentially following the protocols of the SuperScript Plasmid System full details of which are contained in the manufacturer's instruction manual (http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adaptor [5'-pGATGTTTCATGTCGAGCGCGCC(T)15-3']. Not I/blunt end inserts were cloned into the Not I/EcoR V sites in the vector. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

## ORIGIN

Alignment Scores:  
 Pred. No.: 3,986-68 Length: 523  
 Score: 639.00 Matches: 123  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 2 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x BF726890 (1-523)

Qy 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20  
 Db TCCCTCTTCCGACCGCTGCTGACCTCCGGCATCTCTGAGTTCTGATCCGACCGGACGAC 127  
 Qy 21 PheValIlePheLeuAspVallyshisPheSerProGluAspLeuThrVallyshisValGln 40  
 Db TTCGTCATCTTCTCGATGTGAAGCACCTTCTCCCGGAGGACCTCACCGTGAAGGTGCAG 187  
 Qy 41 AspAspPheValGluIleHisGlyIlyshisGluArgGlnAspAspHisGlyTyrIle 60  
 Db GACGACTTTGTGGAGATCCAGGAAGCACACAGAGCGCCAGGACGACGACCGCTACATT 247  
 Qy 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
 Db TCCCGTGAAGTTCACCGCGCTACCGCTGCGCTCCAGTGCAGCAGTGCAGCTCTCTCT 307  
 Qy 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProIlyleIleGlnThrGlyLeu 100  
 Db TGCTCCCTGTCTGCGGATGGCATGCTGACCTTCTGTGCCCCCAAGATCCAGACTGCGCTG 367  
 Qy 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluIlyProThrSerAla 120  
 Db GATGCCACCCAGCCGAGCGACCATCCCGTGTCCGCGGAGGAGAGCCACCTCGGCT 427  
 Qy 121 ProSerSer 123  
 Db CCCCCTGCTC 436

## RESULT 12

BM704974  
 LOCUS 533 bp mRNA linear EST 28-FEB-2002  
 DEFINITION UI-E-DW0-agg-c-09-0-UI.r1 UI-E-DW0 Homo sapiens cDNA clone  
 UI-E-DW0-agg-c-09-0-UI 5', mRNA sequence.  
 ACCESSION BM704974  
 VERSION BM704974.1 GI:19018232  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 533)  
 Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477  
 PUBMED 8899548  
 COMMENT Contact: Soares, MB  
 Coordinated Laboratory for Computational Genomics  
 University of Iowa  
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: bento-soares@uiowa.edu

## COMMENT

Tissue Procurement: Dr. Gregg Hageman  
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research  
 Genetics (www.resgen.com).  
 Seq primer: M13 Reverse.  
 Location/Qualifiers  
 1..533  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-E-DW0-agg-c-09-0-UI"  
 /tissue\_type="lens"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /clone\_lib="UI-E-DW0"  
 /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a  
 modified polylinker; Site 1: EcoR I; Site 2: Not I;  
 UI-E-DW0 is a cDNA library containing the following  
 tissue(s): lens. The library was constructed according to  
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. First strand cDNA synthesis was primed with an  
 oligo-dT primer containing a Not I site. Double stranded  
 cDNA was ligated to an EcoR I adaptor, digested with Not  
 I, and cloned directionally into pT7T3-Pac vector. The  
 oligonucleotide used to prime the synthesis of  
 first-strand cDNA contains a library tag sequence that is  
 located between the Not I site and the (dT)18 tail. The  
 sequence tag for this library is CGATTAGCA. This library  
 was created for the program, Gene Discovery in the Visual  
 System, supported by National Eye Institute (NEI)."

## FEATURES

## source

1..533  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-E-DW0-agg-c-09-0-UI"  
 /tissue\_type="lens"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /clone\_lib="UI-E-DW0"  
 /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a  
 modified polylinker; Site 1: EcoR I; Site 2: Not I;  
 UI-E-DW0 is a cDNA library containing the following  
 tissue(s): lens. The library was constructed according to  
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. First strand cDNA synthesis was primed with an  
 oligo-dT primer containing a Not I site. Double stranded  
 cDNA was ligated to an EcoR I adaptor, digested with Not  
 I, and cloned directionally into pT7T3-Pac vector. The  
 oligonucleotide used to prime the synthesis of  
 first-strand cDNA contains a library tag sequence that is  
 located between the Not I site and the (dT)18 tail. The  
 sequence tag for this library is CGATTAGCA. This library  
 was created for the program, Gene Discovery in the Visual  
 System, supported by National Eye Institute (NEI)."

## ORIGIN

Alignment Scores:  
 Pred. No.: 4,086-68 Length: 533  
 Score: 639.00 Matches: 123  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 4 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x BM704974 (1-533)

Qy 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20  
 Db TCCCTCTTCCGACCGCTGCTGACCTCCGGCATCTCTGAGTTCTGATCCGACCGGACGAC 109  
 Qy 21 PheValIlePheLeuAspVallyshisPheSerProGluAspLeuThrVallyshisValGln 40  
 Db TTCGTCATCTTCTCGATGTGAAGCACCTTCTCCCGGAGGACCTCACCGTGAAGGTGCAG 169  
 Qy 41 AspAspPheValGluIleHisGlyIlyshisGluArgGlnAspAspHisGlyTyrIle 60  
 Db GACGACTTTGTGGAGATCCAGGAAGCACACAGAGCGCCAGGACGACGCTACATT 229  
 Qy 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
 Db TCCCGTGAAGTTCACCGCGCTACCGCTGCGCTCCAGTGCAGCAGTGCAGCTCTCTCT 289

QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProIlyleInThrGlyLeu 100  
 Db 290 TGCTCCCTGTCGCCGATGGATGCTGACCTTCTGTGGCCCAAGATCCAGACTGGCCTG 349  
 QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120  
 Db 350 GATGCCACCCAGCCGAGGAGCCATCCCGGTGTCGGGGAGGAGAGCCACCTCGGCT 409  
 QY 121 ProSerSer 123  
 Db 410 CCTCGTCC 418  
 RESULT 13  
 BM705774  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 534)  
 Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 Genome Res. 6 (9), 791-806 (1996)  
 97044477  
 889548  
 Contact: Soares, MB  
 Coordinated Laboratory for Computational Genomics  
 University of Iowa  
 375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: bento-soares@uiowa.edu  
 Tissue Procurement: Dr. Gregg Hageman  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research  
 Genetics (www.resgen.com).  
 Seq primer: M13 Reverse.  
 Location/Qualifiers  
 1..534  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-E-DW0-agi-h-12-0-UI"  
 /tissue\_type="lens"  
 /dev\_stage="adult"  
 /lab\_host="PH10B (Life Technologies) (T1 phage resistant)"  
 /clone\_lib="UI-E-DW0"  
 /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a  
 modified polylinker; Site\_1: EcoR I; Site\_2: Not I;  
 UI-E-DW0 is a cDNA library containing the following  
 tissue(s): lens. The library was constructed according to  
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. First strand cDNA synthesis was primed with an  
 oligo-dT primer containing a Not I site. Double stranded  
 cDNA was ligated to an EcoR I adaptor, digested with Not  
 I, and cloned directionally into pT7T3-Pac vector. The  
 oligonucleotide used to prime the synthesis of  
 first-strand cDNA contains a library tag sequence that is  
 located between the Not I site and the (dT)18 tail. The  
 sequence tag for this library is CGATTAGCGA. This library  
 was created for the program, Gene Discovery in the Visual  
 System, supported by National Eye Institute (NEI)."

ORIGIN  
 source

Alignment Scores:  
 Pred. No.: 4,096-68 Length: 534  
 Score: 639.00 Matches: 123  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 4 Gaps: 0  
 US-10-657-740-1\_COPY\_51\_173 (1-123) x BM705774 (1-534)  
 QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20  
 Db 78 TCCCTCTTCGCCACCGCTGCTGGACTCCGCGCATCTCTGAGGTTCCATCCGACCGGACAG 137  
 QY 21 PheValIlePheLeuAspValIlyshHisPheSerProGluAspLeuThrValLysValGln 40  
 Db 138 TTCGTGATCTTCTCGATGTGAAGCACTTCTCCCGGAGGACCTCACCGTGAAGGTGCAG 197  
 QY 41 AspAspPheValGluIleHisGlyIlyshHisAsnGluArgGlnAspAspHisGlyIlyle 60  
 Db 198 GACGACTTTGTGGAGATCCACGGAAGCACAAACGAGCGCCAGGACGACCGGTACATT 257  
 QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
 Db 258 TCCCGTGAATTCCACCGCCGCTACCGCTGCGCTCAACGTGGACCACTGGCCCTCTCT 317  
 QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProIlyleGlnThrGlyLeu 100  
 Db 318 TGCTCCCTGTCGCGATGGCATCTGACCTTCTGTGGCCCCCAAGATCCAGACTGGCCTG 377  
 QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluLysProThrSerAla 120  
 Db 378 GATCCACCCACCGCCGAGGACCATCCCGTGTGCGGGAGGAGAGCCACCTCGGCT 437  
 QY 121 ProSerSer 123  
 Db 438 CCTCGTCC 446  
 RESULT 14  
 BM726399  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 536)  
 Wistow,G.J., Bernstein,S., Behal,A. and Smith,D.  
 NEIBANK: EST analysis and bioinformatics for ocular genomics  
 Invest. Ophthalmol. Vis. Sci. 41 (2000) In press  
 Contact: Wistow G  
 Section on Molecular Structure and Function  
 National Eye Institute  
 6/331, NIH, Bethesda, MD 20892-2740, USA  
 Tel: 301 402 3452  
 Fax: 301 496 0078  
 Email: graeme@helix.nih.gov  
 Plate: 06 row: a column: 09  
 Seq primer: M13Rp1 reverse primer (ABI).  
 Location/Qualifiers  
 1..536  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="by06a09"  
 /tissue\_type="Lens"  
 /dev\_stage="Adult"  
 /lab\_host="EMDH10B"  
 /clone\_lib="Human Lens cDNA (Un-normalized, unamplified):"

FEATURES  
 source



BY"

/note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses from different adults (both approximately 40 years old) together yielded 20ug of total RNA and 150ng mRNA for cDNA library synthesis. A directionally cloned cDNA library in the pCMVSPORT6 vector was constructed at Life Technologies, essentially following the protocols of the SuperScript Plasmid System full details of which are contained in the manufacturer's instruction manual (http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adaptor [5'-pGACTAGTCTAGATCGGAGCGGCC(T)15-3']. Not I/blunt end inserts were cloned into the Not I/EcoR V sites in the vector. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

## ORIGIN

Alignment Scores:  
 Pred. No.: 4,11e-68 Length: 536  
 Score: 639.00 Matches: 123  
 Percent Similarity: 100.00% Conservativeness: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 2 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x BF726399 (1-536)

QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20  
 DB 162 TCCCTCTTCCGACCGCTGCTGAGCTCCGGCATCTCTGAGGTTCCGATCCGCGGACAG 221  
 QY 21 PheValIlePheLeuAspVallyshisPheSerProGluAspLeuThrVallyshisValGln 40  
 DB 222 TTCGTTCATCTTCTCCTGATGTGAAGCATTCTCCCGGAGGACCTCACCGTGAAGGTGCAG 281  
 QY 41 AspAspPheValGluIleHisGlyIleHisGluArgGlnAspAspHisGlyIleTyrIle 60  
 DB 282 GACGACTTTGTGGAGATCCAGGAAGCACACGAGCGCCAGGACGACCGCGGTACATT 341  
 QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
 DB 342 TCCCGTGAGTTCACCGNGCTACCGCTCGCTCCAGTGGACATCGGCGCTCTCT 401  
 QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLyseIleGlnThrGlyLeu 100  
 DB 402 TGCTCCCTGTCTCCGATGGATGCTGACCTTCTGTGCCCCAGATCCAGACTGGCCTG 461  
 QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120  
 DB 462 GATGCCACCCAGCCGCGAGGAGCCATCCCGTGTCCGCGGAGGAGAGCCACCTCGGCT 521  
 QY 121 ProSerSer 123  
 DB 522 CCCTCGTCC 530

## RESULT 15

BF726679  
 LOCUS BF726679 558 bp mRNA linear EST 05-JAN-2001  
 DEFINITION by10d02.y1 Human Lens cDNA (Un-normalized, unamplified): BY Homo sapiens cDNA clone by10d02 5', mRNA sequence.

ACCESSION BF726679

VERSION BF726679.1 GI:12042590

KEYWORDS EST.

SOURCE Homo sapiens (human)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 558)

Wistow G.J., Bernstein S., Behal A. and Smith, D.

NEIBANK: EST analysis and bioinformatics for ocular genomics

Invest. Ophthalmol. Vis. Sci. 41 (2000) In press

Contact: Wistow G

Section on Molecular Structure and Function

National Eye Institute  
 6/331, NIH, Bethesda, MD 20892-2740, USA  
 Tel: 301 402 3452  
 Fax: 301 496 0078  
 Email: graeme@helix.nih.gov  
 Plate: 10 row: d column: 02  
 Seq primer: M13RP1 reverse primer (ABI).

## FEATURES

source  
 1..558  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="by10d02"  
 /tissue\_type="Lens"  
 /dev\_stage="Adult"  
 /lab\_host="EMDH10B"  
 /clone\_lib="Human Lens cDNA (Un-normalized, unamplified): BY"

/note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses from different adults (both approximately 40 years old) together yielded 20ug of total RNA and 150ng mRNA for cDNA library synthesis. A directionally cloned cDNA library in the pCMVSPORT6 vector was constructed at Life Technologies, essentially following the protocols of the SuperScript Plasmid System full details of which are contained in the manufacturer's instruction manual (http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adaptor [5'-pGACTAGTCTAGATCGGAGCGGCC(T)15-3']. Not I/blunt end inserts were cloned into the Not I/EcoR V sites in the vector. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

## ORIGIN

Alignment Scores:  
 Pred. No.: 4,33e-68 Length: 558  
 Score: 639.00 Matches: 123  
 Percent Similarity: 100.00% Conservativeness: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 2 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x BF726679 (1-558)

QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20  
 DB 140 TCCCTCTTCCGACCGCTGCTGAGCTCCGGCATCTCTGAGGTTCCGATCCGCGGACAG 199  
 QY 21 PheValIlePheLeuAspVallyshisPheSerProGluAspLeuThrVallyshisValGln 40  
 DB 200 TTCGTTCATCTTCTCCTGATGTGAAGCATTCTCCCGGAGGACCTCACCGTGAAGGTGCAG 259  
 QY 41 AspAspPheValGluIleHisGlyIleHisGluArgGlnAspAspHisGlyIleTyrIle 60  
 DB 260 GACGACTTTGTGGAGATCCAGGAAGCACACGAGCGCCAGGACGACCGCGGTACATT 319  
 QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
 DB 320 TCCCGTGAGTTCACCGCGCTACCGCTCGCTCCAGCTGAGGACGACCTCGGCGCTCTCT 379  
 QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLyseIleGlnThrGlyLeu 100  
 DB 380 TGCTCCCTGTCTCCGATGGATGCTGACCTTCTGTGCCCCAGATCCAGACTGGCCTG 439  
 QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120  
 DB 440 GATGCCACCCAGCCGCGAGGAGCCATCCCGTGTCCGCGGAGGAGAGCCACCTCGGCT 499  
 QY 121 ProSerSer 123  
 DB 500 CCCTCGTCC 508

## RESULT 16

BM697368 569 bp mRNA linear EST 28-FEB-2002  
LOCUS UI-E-DW0-agm-d-06-0-UI.r1 UI-E-DW0 Homo sapiens cDNA clone  
DEFINITION UI-E-DW0-agm-d-06-0-UI 5', mRNA sequence.  
ACCESSION BM697368.1 GI:19010626  
VERSION EST.  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 569)  
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene  
discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477  
PubMed 889548  
COMMENT Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@uiowa.edu  
Tissue Procurement: Dr. Gregg Hageman  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics (www.resgen.com).  
Seq primer: M13 Reverse.  
Location/Qualifiers  
1. .569  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-E-DW0-agm-d-06-0-UI"  
/tissue\_type="lens"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone\_lib="UI-E-DW0"  
/notes="Organ: eye; Vector: pT73-Pac (Pharmacia) with a  
modified polylinker; Site 1: EcoR I; Site 2: Not I;  
UI-E-DW0 is a cDNA library containing the following  
tissue(s): lens. The library was constructed according to  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. First strand cDNA synthesis was primed with an  
oligo-dT primer containing a Not I site. Double stranded  
cDNA was ligated to an EcoR I adaptor, digested with Not  
I, and cloned directionally into pT73-Pac vector. The  
oligonucleotide used to prime the synthesis of  
first-strand cDNA contains a library tag sequence that is  
located between the Not I site and the (dr)18 tail. The  
sequence tag for this library is CGATTAGCGA. This library  
was created for the program, Gene Discovery in the Visual  
System, supported by National Eye Institute (NEI)."

Alignment Scores:  
Pred. No.: 4 44e-68 Length: 569  
Score: 639.00 Matches: 123  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 4 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x BM697368 (1-569)

QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspIys 20  
Db 143 TCCCTCTTCCGACCGCTCGACTCCGGCATCTCTGAGGTTCGATCCGACCGGACAAG 202

QY 21 PheValIlePheLeuAspValIysHisPheSerProGluAspLeuThrValIysValGln 40  
Db 203 TTTCGTCACTTCTCTCGATGTGAAGCACCTTCTCCCGGAGGAGCTCACCGTGAAGGTGCAG 262  
QY 41 AspAspPheValGluIleHisGlyIysHisGlnGluArgGlnAspAspHisGlyTyrIle 60  
Db 263 GACGACTTTGTGGAGATCCAGGAAGCACACACGAGCGCCAGGACGACCGCTTACATT 322  
QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
Db 323 TCCCGTGAAGTCCACCGCGCTACCGCTCGCTCAACGTCGAGGACGTCGTGGCCCTCTCT 382  
QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100  
Db 383 TGCTCCCTGTCTGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 442  
QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120  
Db 443 GATGCCACCCACGCGAGCGAGCCATCCCGTGTGCGGGAGGAGAGAGCCACCTCGCT 502  
QY 121 ProSerSer 123  
Db 503 CCTCGTCC 511  
RESULT 17  
BF726236  
LOCUS  
DEFINITION by03f01.y1 Human Lens cDNA (Un-normalized, unamplified): BY Homo  
sapiens cDNA clone by03f01 5', mRNA sequence.  
ACCESSION BF726236  
VERSION BF726236.1 GI:12042147  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 576)  
AUTHORS Wistow,G.J., Bernstein,S., Behal,A. and Smith,D.  
TITLE NEIBANK: EST analysis and bioinformatics for ocular genomics  
JOURNAL Invest. Ophthalmol. Vis. Sci. 41 (2000) In press  
COMMENT Contact: Wistow G  
Section on Molecular Structure and Function  
National Eye Institute  
6/331, NIH, Bethesda, MD 20892-2740, USA  
Tel: 301 402 3452  
Fax: 301 496 0078  
Email: graeme@helix.nih.gov  
Plate: 03 row: f column: 01  
Seq primer: M13RP1 reverse primer (ABI).  
Location/Qualifiers  
1. .576  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="by03f01"  
/tissue\_type="lens"  
/dev\_stage="Adult"  
/lab\_host="EMDH10B"  
/clone\_lib="Human Lens cDNA (Un-normalized, unamplified):  
BY"

/notes="Organ: Eye; Vector: pCMVSPORT6; Two human lenses  
from different adults (both approximately 40 years old)  
together yielded 20ug of total RNA and 150ng mRNA for cDNA  
library synthesis. A directionally cloned cDNA library in  
the pCMVSPORT6 vector was constructed at Life  
Technologies, essentially following the protocols of the  
SuperScript Plasmid System full details of which are  
contained in the manufacturer's instruction manual  
(http://www.lifetech.com/). First strand synthesis was  
carried out using a Not I primer-adaptor  
[5'-pGACTAGTTCAGATCGGAGCGGCCCT(T)15-3']. Not I/blunt  
ends were cloned into the Not I/EcoR V sites in the  
vector. EST analysis was performed on the unamplified

library at the NIH Intramural Sequencing Center (NISC)."

## ORIGIN

## Alignment Scores:

Pred. No.: 4,51e-68 Length: 576  
Score: 639.00 Matches: 123  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 2 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x BF726236 (1-576)

QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20  
DB 198 TCCCTCTCCGACCGCTGCGACCTCCGCATCTCTGAGGTTGATCCGACCGGCAAG 257  
QY 21 PheValIlePheLeuAspVallyHisPheSerProGluAspLeuThrVallyValGln 40  
DB 258 TTCGTATCTTCTCGATGTGAAGCACTTCTCCCGAGGACCTCACCGTGAAGTGCAG 317  
QY 41 AspAspPheValGluIleHisGlyIleHisAsnGluArgGlnAspHisGlyTyrIle 60  
DB 318 GACGACTTTGTGAGATCCACGGAAGCAACACGAGCGCCAGCAGCACCGCTACATT 377  
QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
DB 378 TCCCGTGAAGTTCCACCGCGCTACCGCTCCGCTCCGCTCCGAGGACCTCCGCTCTCT 437  
QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProlystleGlnThrGlyLeu 100  
DB 438 TGCTCCCTGTGCGCATGCGATGCTGACCTTCTGTGGCCCCAAGATCCAGACTGCGCTG 497  
QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120  
DB 498 GATGCCACCCACCGCGAGGAGCCATCCCGTGTGCGGGGAGGAGAGGCCACCTCGCT 557  
QY 121 ProSerSer 123  
DB 558 CCTCGTCC 566

## RESULT 18

BF727002 577 bp mRNA linear EST 05-JAN-2001  
LOCUS by15c11.y1 Human Lens cDNA (Un-normalized, unamplified): BY Homo  
DEFINITION sapiens cDNA clone by15c11 5', mRNA sequence.

ACCESSION BF727002

VERSION BF727002.1 GI:12042913

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 577)

WISOW,G.J., Bernstein,S., Behal,A. and Smith,D.

NEIBANK: EST analysis and bioinformatics for ocular genomics

Invest. Ophthalmol. Vis. Sci. 41 (2000) In press

Contact: Wistow G

Section on Molecular Structure and Function

National Eye Institute

6/331, NIH, Bethesda, MD 20892-2740, USA

Tel: 301 402 3452

Fax: 301 496 0078

Email: graeme@helix.nih.gov

Plate: 15 row: C column: 11

Seq primer: M13RPI reverse primer (ABI).

Location/Qualifiers

1..577

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="by15c11"

/tissue\_type="Lens"

## FEATURES

source

BM721893 578 bp mRNA linear EST 01-MAR-2002  
LOCUS UI-E-E00-ahw-c-15-0-UI.r1 UI-E-E00 Homo sapiens cDNA clone  
DEFINITION UI-E-E00-ahw-c-15-0-UI 5', mRNA sequence.

ACCESSION BM721893

VERSION BM721893.1 GI:19042101

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 578)

Bonaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene

/dev\_stage="Adult"  
/lab\_host="EMDH10B"  
/clone\_lib="Human Lens cDNA (Un-normalized, unamplified):  
BY"

/note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses  
from different adults (both approximately 40 years old)  
together yielded 20ug of total RNA and 150ng mRNA for cDNA  
library synthesis. A directionally cloned cDNA library in  
the pCMVSPORT6 vector was constructed at Life  
Technologies, essentially following the protocols of the  
SuperScript Plasmid System full details of which are  
contained in the manufacturer's instruction manual  
(http://www.lifetech.com/). First strand synthesis was  
carried out using a Not I primer-adaptor  
[5'-pGACTAGTCTAGATCGGAGCGGCCCT(T)15-3']. Not I/blunt  
ends were cloned into the Not I/EcoR V sites in the  
vector. EST analysis was performed on the unamplified  
library at the NIH Intramural Sequencing Center (NISC)."

## ORIGIN

## Alignment Scores:

Pred. No.: 4,52e-68 Length: 577  
Score: 639.00 Matches: 123  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 2 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x BF727002 (1-577)

QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20  
DB 145 TCCCTCTTCCGACCGCTGCTGACCTCCGCATCTCTGAGGTTGATCCGACCGGCAAG 204  
QY 21 PheValIlePheLeuAspVallyHisPheSerProGluAspLeuThrVallyValGln 40  
DB 205 TTCGTATCTTCTCGATGTGAAGCACTTCTCCCGAGGACCTCACCGTGAAGTGCAG 264  
QY 41 AspAspPheValGluIleHisGlyIleHisAsnGluArgGlnAspHisGlyTyrIle 60  
DB 265 GACGACTTTGTGAGATCCACGGAAGCAACACGAGCGCCAGCAGCACCGCTACATT 324  
QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
DB 325 TCCCGTGAAGTTCCACCGCGCTACCGCTCCGCTCCGCTCCGAGGACCTCCGCTCTCT 384  
QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProlystleGlnThrGlyLeu 100  
DB 385 TGCTCCCTGTGCGCATGCGATGCTGACCTTCTGTGGCCCCAAGATCCAGACTGCGCTG 444  
QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120  
DB 445 GATGCCACCCACCGCGAGGAGCCATCCCGTGTGCGGGGAGGAGAGGCCACCTCGCT 504  
QY 121 ProSerSer 123  
DB 505 CCTCGTCC 513

## RESULT 19

BM721893

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 578)

Bonaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene

JOURNAL  
MEDLINE  
PUBMED  
COMMENT

discovery  
Genome Res. 6 (9), 791-806 (1996)  
97044477  
8889548  
Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@uiowa.edu  
Tissue Procurement: Dr. Gregg Hageman  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
cDNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics (www.reagen.com).  
Seq primer: M13 Reverse.

FEATURES  
source

Location/Qualifiers  
1..578  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-E-E00-ahw-c-15-0-UI"  
/tissue\_type="fetal eye"  
/dev\_stage="fetal"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone\_lib="UI-E-E00"  
/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-E00 is a cDNA library containing the following tissue(s): fetal eye. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CGCGTATACC. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

ORIGIN

Alignment Scores:  
Pred. No.: 4,53e-68 Length: 578  
Score: 639.00 Matches: 123  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 4 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x BM721893 (1-578)

QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20  
205 TCCTCTTCGACACCGTGTGACTCCGGCATCTCTGAGGTTGATCCGACCGGACAAG 264  
QY 21 PheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrValLysValGln 40  
265 TTGCTCATCTTCCTCGATGTGAAGCACTTCTCCCGGAGGACCTCACCGTGNAGTGCAG 324  
QY 41 AspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspHisGlyTyrIle 60  
325 GACGACTTTGTGAGATCCACGAAGACACACGAGCGCCAGCAGCAGCGGTACATT 384  
QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
385 TCCCGTGAGTTCCACCGCGCTACCGCTGCGCTCCACGTGGACCATCGGCTCTCT 444  
QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100

Db 445 TGCTCCCTGTCTGCCGATGGCATGCTGACCTTCTGTGGCCCCCAAGATCCAGACTGCGCTG 504  
QY 101 AspAlaThrHisAlaGluAlaGluAlaIleProValSerArgGluGluLysProThrSerAla 120  
|||||  
Db 505 GATGCCACCCAGCCCGGAGCGCCATCCCGTGTGCGGGAGGAGAGCCACCTCGGCT 564  
QY 121 ProSerSer 123  
|||||  
Db 565 CCCTCGTCC 573

RESULT 20  
BF727295

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BF727295 592 bp mRNA linear EST 05-JAN-2001  
by19e10.y1 Human Lens cDNA (Un-normalized, unamplified): BY Homo  
sapiens cDNA clone by19e10 5', mRNA sequence.

BF727295  
BF727295.1 GI:12043206

EST.  
Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Wistow,G.J., Bernstein,S., Behal,A. and Smith,D.  
NEIBANK: EST analysis and bioinformatics for ocular genomics

Invest. Ophthalmol. Vis. Sci. 41 (2000) In press

Contact: Wistow G

Section on Molecular Structure and Function

National Eye Institute

6/331, NIH, Bethesda, MD 20892-2740, USA

Tel: 301 402 3452

Fax: 301 496 0078

Email: graeme@helix.nih.gov

Plate: 19 row: e column: 10

Seq primer: M13RP1 reverse primer (ABI).

Location/Qualifiers

1..592

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="by19e10"

/tissue\_type="Lens"

/dev\_stage="Adult"

/lab\_host="EMDH10B"

/clone\_lib="Human Lens cDNA (Un-normalized, unamplified):

BY"

/note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses

from different adults (both approximately 40 years old)

together yielded 20ug of total RNA and 150ng mRNA for cDNA

library synthesis. A directionally cloned cDNA library in

the pCMVSPORT6 vector was constructed at Life

Technologies, essentially following the protocols of the

SuperScript Plasmid System full details of which are

contained in the manufacturer's instruction manual

(http://www.lifetech.com/). First strand synthesis was

carried out using a Not I primer-adaptor

[5'-pCAGTAGTTCTAGATCGGAGCGCGCC(T)15-3']. Not I/blunt

end inserts were cloned into the Not I/EcoR V sites in the

vector. EST analysis was performed on the unamplified

library at the NIH Intramural Sequencing Center (NISC).

ORIGIN

Alignment Scores:

Pred. No.: 4.68e-68 Length: 592  
Score: 639.00 Matches: 123  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 2 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x BF727295 (1-592)

QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20

```

Db      161 TCCCTCTCCGACCGTGTGGACTCCGGCATCTTGAGGTTCGATCCGACCGGACAAG 220
QY      21  PheValIlePheLeuAspVallyHisPheSerProGluAspLeuThrVallyValGln 40
Db      221 TTCTGTCATCTCTCGATGTGAAGCACTTCTCCCGGAGGACCTCAACCGTGAAGGTGCAG 280
QY      41  AspAspPheValIleHisGlyHisAsnGluArgGlnAspAspHisGlyTyrIle 60
Db      281 GACGACTTTGTGGAGATCCACGAAAGCAACAGCAGCGCCAGACACCGCTACAT 340
QY      61  SerArgGluPheHisArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80
Db      341 TCCCGTGAATTCACCGCGCTACCGCTGCGCTGCTCAACGTGAGACCGTCCCTCTCT 400
QY      81  CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100
Db      401 TGCTCCCTGCTGTCCGATGCGATGCTGACCTTCTGTGGCCCCCAAGATCCAGACTGCGCTG 460
QY      101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120
Db      461 GATGCCACCCAGCGGAGCGGACCATCCCGTGTGCGGGAGGAGAAGCCACCTCGGCT 520
QY      121 ProSerSer 123
Db      521 CCCTCGTCC 529

BM705926 592 bp mRNA linear EST 28-FEB-2002
UI-E-DW0-agh-f-18-0-UI.r1 UI-E-DW0 Homo sapiens cDNA clone
UI-E-DW0-agh-f-18-0-UI 5', mRNA sequence.
BM705926
BM705926.1 GI:19019184
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 592)
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
8889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA library prepared by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 Reverse.
Location/Qualifiers
1..592
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-DW0-agh-f-18-0-UI"
/tissue_type="lens"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: p773-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-E-DW0 is a cDNA library containing the following

```

tissue(s): lens. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CGATTAGCGA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

## ORIGIN

```

Alignment Scores:
Pred. No.: 4,68e-68 Length: 592
Score: 639.00 Matches: 123
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

```

US-10-657-740-1\_COPY\_51\_173 (1-123) x BM705926 (1-592)

```

QY      1  SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20
Db      202 TCCCTCTTCCGACCGTGTGGACTCCGGCATCTCTGAGGTTCCGATCCGACCGGACAAG 261
QY      21  PheValIlePheLeuAspVallyHisPheSerProGluAspLeuThrVallyValGln 40
Db      262 TTCTGTCATCTCTCGATGTGAAGCACTTCTCCCGGAGGACCTCAACCGTGAAGGTGCAG 321
QY      41  AspAspPheValIleHisGlyHisAsnGluArgGlnAspAspHisGlyTyrIle 60
Db      322 GACGACTTTGTGAGATCCACGAAAGCAACAGCAGCGCCAGACACCGCTACAT 381
QY      61  SerArgGluPheHisArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80
Db      382 TCCCGTGAATTCACCGCGCTACCGCTGCGCTGCTCAACGTGAGGAGGAGAAGCCACCTCG 441
QY      81  CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100
Db      442 TGCTCCCTGCTGTCCGATGCGATGCTGACCTTCTGTGGCCCCCAAGATCCAGACTGCGCTG 501
QY      101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120
Db      502 GATGCCACCCAGCGGAGCGGACCATCCCGTGTGCGGGAGGAGAAGCCACCTCGGCT 561
QY      121 ProSerSer 123
Db      562 CCCTCGTCC 570

```

## RESULT 22

## BM697101

## LOCUS

## DEFINITION

## UI-E-DW0-agh-m-13-0-UI.r1 UI-E-DW0 Homo sapiens cDNA clone

## UI-E-DW0-agh-m-13-0-UI 5', mRNA sequence.

## ACCESSION

## BM697101

## VERSION

## BM697101.1

## KEYWORDS

## EST.

## SOURCE

## Homo sapiens (human)

## ORGANISM

## Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

## Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## 1 (bases 1 to 593)

## Normalization and subtraction: two approaches to facilitate gene

## discovery

## Genome Res. 6 (9), 791-806 (1996)

## JOURNAL

## MEDLINE

## PUBMED

## COMMENT

## Contact: Soares, MB

## Coordinated Laboratory for Computational Genomics

University of Iowa  
375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@uiowa.edu

Tissue Procurement: Dr. Gregg Hageman  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).  
Seq primer: M13 Reverse.

## FEATURES

source

Location/Qualifiers  
1..596  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-E-DW0-aggm-m-13-0-UI"  
/dev\_stage="adult"  
/tissue\_type="lens"  
/lab\_host="DHI0B (Life Technologies) (T1 phage resistant)"  
/clone\_lib="UI-E-DW0"  
/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoR I; Site\_2: Not I; UI-E-DW0 is a cDNA library containing the following tissue(s): lens. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CGATTAGCGA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

## ORIGIN

## Alignment Scores:

Pred. No.: 4,696-68 Length: 593  
Score: 639.00 Matches: 123  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 4 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x BM697101 (1-593)

Qy	1	SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys	20
Db	136	TCCTCTTCGCGACCGTGTGACTCCGGCACTCTGAGGTTCGATCCGCGGACAA	195
Qy	21	PheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrValLysValGln	40
Db	196	TTGCTCATCTTCCTGATGTGAAGCACTTCTCCCGAGGACCTCACCCTGAGGTGCAG	255
Qy	41	AspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIle	60
Db	256	GACGACTTGTGGAGATCCAGGAAGCACACGAGCGCCAGGACGACACCGGTACATT	315
Qy	61	SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer	80
Db	316	TCCCGTGTAGTTCCACGCGCGCTACCGCTGCGTCCACGTGGACCACTGCGCCCTCTCT	375
Qy	81	CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu	100
Db	376	TGCTTCCTGTTCGCGATGGATGCTGACCTTCTGTGGCCCCAAGATCCAGCTGGCGTG	435
Qy	101	AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla	120
Db	436	GATGCCACCCACCGCGAGGCCATCCCTGTTCGCGGGAGGAGAAACCCCTCGGCT	495

Qy 121 ProSerSer 123  
Db 496 CCCTCGTCC 504

## RESULT 23

BM696581

LOCUS

DEFINITION

UI-E-DW0-agg-o-15-0-UI.r1 UI-E-DW0 Homo sapiens cDNA clone  
UI-E-DW0-agg-o-15-0-UI 5', mRNA sequence.

ACCESSION

BM696581

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 596)

AUTHORS

Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE

Normalization and subtraction: two approaches to facilitate gene

discovery

JOURNAL

Genome Res. 6 (9), 791-806 (1996)

MEDLINE

97044477

PUBMED

8889548

COMMENT

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

Tissue Procurement: Dr. Gregg Hageman

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

Clone Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research

Genetics (www.resgen.com).

Seq primer: M13 Reverse.

Location/Qualifiers

1..596

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="UI-E-DW0-agg-o-15-0-UI"

/tissue\_type="lens"

/dev\_stage="adult"

/lab\_host="DHI0B (Life Technologies) (T1 phage resistant)"

/clone\_lib="UI-E-DW0"

/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoR I; Site\_2: Not I; UI-E-DW0 is a cDNA library containing the following tissue(s): lens. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CGATTAGCGA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

## ORIGIN

## Alignment Scores:

Pred. No.: 4,726-68 Length: 596  
Score: 639.00 Matches: 123  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 4 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x BM696581 (1-596)

QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20  
 Db 206 TCCCTCTTCCGACCGCTGCGCATCTCTGAGTTCTGATCCGACCGGACCAAG 265  
 QY 21 PheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrValLysValGln 40  
 Db 266 TTGCTCACTTCTCGATGTGAGCACTTCTCCCGGAGGACCTCACCGTGAAGTGGAG 325  
 QY 41 AspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIle 60  
 Db 326 GACGACTTGTGAGATCCACGGAAGCACACGAGCGCCAGGACGACCCACGCTTACATT 385  
 QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
 Db 386 TCCCGTGAAGTTCCACCGCGCTACCGCTCGCGTCCACGTCGAGCAGCTCGCGCTCTCT 445  
 QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100  
 Db 446 TGCTCCCTGTGCGCATGTGCGATGCTGACCTTCTGTGGCCCCAAGATCCAGACTGGCTG 505  
 QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120  
 Db 506 GATGCCACCCAGCGCGAGGACCATCCCGTGTGCGGGAGGAGAGCCACCTCGGCT 565  
 QY 121 ProSerSer 123  
 Db 566 CCCTCGTCC 574

RESULT 24  
 BM696976 599 bp mRNA linear EST 28-FEB-2002  
 LOCUS UI-E-DW0-agl-f-04-0-UI.r1 UI-E-DW0 Homo sapiens cDNA clone  
 DEFINITION UI-E-DW0-agl-f-04-0-UI 5', mRNA sequence.  
 ACCESSION BM696976  
 VERSION BM696976.1 GI:19010234  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 599)  
 REFERENCE Ronaldo,M.F., Lennon,G. and Soares,M.B.  
 AUTHORS Normalization and subtraction: two approaches to facilitate gene  
 TITLE discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 9704477  
 PUBMED 889548  
 COMMENT Contact: Soares, MB  
 Coordinated Laboratory for Computational Genomics  
 University of Iowa  
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: bento-soares@uiowa.edu  
 Tissue Procurement: Dr. Gregg Hageman  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research  
 Genetics (www.regen.com).  
 Seq primer: M13 Reverse.  
 Location/Qualifiers  
 1: 599  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-E-DW0-agl-f-04-0-UI"  
 /tissue\_type="lens"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /clone\_lib="UI-E-DW0"  
 /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a

modified polylinker; Site 1: EcoR I; Site 2: Not I;  
 UI-E-DW0 is a cDNA library containing the following  
 tissue(s): lens. The library was constructed according to  
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. First strand cDNA synthesis was primed with an  
 oligo-dT primer containing a Not I site. Double stranded  
 cDNA was ligated to an EcoR I adaptor, digested with Not  
 I, and cloned directionally into pT7T3-Pac vector. The  
 oligonucleotide used to prime the synthesis of  
 first-strand cDNA contains a library tag sequence that is  
 located between the Not I site and the (dT)18 tail. The  
 sequence tag for this library is CGATTAGCGA. This library  
 was created for the program, Gene Discovery in the Visual  
 System, supported by National Eye Institute (NEI)."

## ORIGIN

Alignment Scores:  
 Pred. No.: 4,75e-68 Length: 599  
 Score: 639.00 Matches: 123  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 4 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x BM696976 (1-599)

QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20  
 Db 70 TCCCTCTTCCGACCGCTGCGCATCTCTGAGTTCTGATCCGACCGGACCAAG 129  
 QY 21 PheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrValLysValGln 40  
 Db 130 TTGCTCACTTCTCGATGTGAGCACTTCTCCCGGAGGACCTCACCGTGAAGTGGCAG 189  
 QY 41 AspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIle 60  
 Db 190 GACGACTTGTGAGATCCACGGAAGCACACGAGCGCCAGGACGACCCACGCTTACATT 249  
 QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
 Db 250 TCCCGTGAAGTTCCACCGCGCTACCGCTCGCGTCCACGTCGAGCAGTCGCGCTCTCT 309  
 QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100  
 Db 310 TGCTCCCTGTGTCGCGATGGCATGTGTCACCTTCTGTGGCCCCAGATCCAGACTGCGCTG 369  
 QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120  
 Db 370 GATGCCACCCAGCGGAGGACCATCCCGTGTGCGGGAGGAGAGCCACCTCGGCT 429  
 QY 121 ProSerSer 123  
 Db 430 CCCTCGTCC 438

## RESULT 25

CD675250 607 bp mRNA linear EST 24-JUN-2003  
 LOCUS fs21c02.v1 Human Lens cDNA (Normalized): fs Homo sapiens cDNA clone  
 DEFINITION fs21c02.5', mRNA sequence.  
 ACCESSION CD675250  
 VERSION CD675250.1 GI:32176981  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 607)  
 REFERENCE Wistow,G., Bernstein,S.I., Wyatt,M.K., Behal,A., Touchman,J.W.,  
 AUTHORS Bouffard,G., Smith,D. and Peterson,K.  
 TITLE Expressed sequence tag analysis of adult human lens for the NEIBank  
 Project: over 2000 non-redundant transcripts, novel genes and  
 splice variants  
 JOURNAL Mol. Vis. 8 (4), 171-184 (2002)

MEDLINE  
PUBMED  
COMMENT

22103463  
12107413  
Contact: Wistow G  
Section on Molecular Structure and Function  
National Eye Institute  
6/331, NIH, Bethesda, MD 20892-2740, USA  
Tel: 301 402 3452  
Fax: 301 496 0078  
Email: graeme@helix.nih.gov  
Plate: 21 row: c column: 02  
Seq primer: M13RPI reverse primer (ABI).  
Location/Qualifiers  
1. .607  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="fs21c02"  
/tissue\_type="lens"  
/dev\_stage="Adult"  
/lab\_host="EMDH10B"  
/clone\_lib="Human Lens cDNA (Normalized): fs"  
/note="Organ: Eye; Vector: pCMVSPORT6; A human lens library (by) was normalized by self-subtraction. One portion of double stranded plasmid DNA representing the library was linearized by NotI. This NotI digested library was used as a template for biotinylated RNA synthesis using SP6 RNA polymerase. Another portion of the double stranded plasmid library was converted to single-stranded circles in vitro using Gene II and Exonuclease III (Life Technologies). Single-stranded DNA (1 mg) was hybridized (Cot 500) with 41 mg of Bio-RNA and vector blocking oligonucleotides. The hybridized Bio-RNA/ss-circles were removed by streptavidin:phenol extraction. EST analysis was performed on the library at the NIH Intramural Sequencing Center (NISC)."

## FEATURES

source

Db 561 CCCTCGTCC 569

RESULT 26  
LOCUS CD676130  
DEFINITION fs32e02.y1 Human Lens cDNA (Normalized): fs Homo sapiens CDNA clone fs32e02 5', mRNA sequence.  
ACCESSION CD676130  
VERSION CD676130  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 612)  
AUTHORS Wistow,G., Bernstein,S.L., Wyatt,M.K., Behal,A., Touchman,J.W., Bouffard,G., Smith,D. and Peterson,K.  
TITLE Expressed sequence tag analysis of adult human lens for the NEIBank Project: over 2000 non-redundant transcripts, novel genes and splice variants  
JOURNAL Mol. Vis. 8 (4), 171-184 (2002)  
MEDLINE 22103463  
PUBMED 12107413  
COMMENT Contact: Wistow G  
Section on Molecular Structure and Function  
National Eye Institute  
6/331, NIH, Bethesda, MD 20892-2740, USA  
Tel: 301 402 3452  
Fax: 301 496 0078  
Email: graeme@helix.nih.gov  
Plate: 32 row: e column: 02  
Seq primer: M13RPI reverse primer (ABI).  
Location/Qualifiers  
1. .612  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="fs32e02"  
/tissue\_type="lens"  
/dev\_stage="Adult"  
/lab\_host="EMDH10B"  
/clone\_lib="Human Lens cDNA (Normalized): fs"  
/note="Organ: Eye; Vector: pCMVSPORT6; A human lens library (by) was normalized by self-subtraction. One portion of double stranded plasmid DNA representing the library was linearized by NotI. This NotI digested library was used as a template for biotinylated RNA synthesis using SP6 RNA polymerase. Another portion of the double stranded plasmid library was converted to single-stranded circles in vitro using Gene II and Exonuclease III (Life Technologies). Single-stranded DNA (1 mg) was hybridized (Cot 500) with 41 mg of Bio-RNA and vector blocking oligonucleotides. The hybridized Bio-RNA/ss-circles were removed by streptavidin:phenol extraction. EST analysis was performed on the library at the NIH Intramural Sequencing Center (NISC)."

## FEATURES

source

Alignment Scores:  
Pred. No.: 4,88e-68 Length: 607  
Score: 639.00 Matches: 123  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x CD675250 (1-607)

QY 1 SerLeupheArgThrValLeuAspSerGlyIleSerGluValArgSerAspAspAsp 20  
Db 201 TCCTCTTCCGACCGTGTGACTCCGACATCTCTGAGTTTCGATCCGACCGGACGACG 260  
QY 21 PheValIlePheLeuAspVallyshisPheSerProGluAspLeuThrVallyValGln 40  
Db 261 TTCGTATCTTCTCGATGTGAAGACATCTTCCCGGAGGACCTCACCCTGAAGGTGCAG 320  
QY 41 AspAspPheValGluIleHisGlyIlyshisAsnGluArgGlnAspAspHisGlyTyrIle 60  
Db 321 GACGACTTTGTGAGATCCAGGAAGCACACGAGCCGACGACGACGACGACGACGACG 380  
QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
Db 381 TCCCGTGAGTTCCACCGCGGTACCGCTGCGCTCCAACTGGACGACGACGACGACGACG 440  
QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProIlyshisGlnThrGlyLeu 100  
Db 441 TGCTCTCTGTCTGCGATGCGATGCTGACCTTCTGTGGCCCCAAGATCCAGACTGGCCTG 500  
QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluIlyshisProThrSerAla 120  
Db 501 GATGCCACCCACGCGAGCGACCATCCCGTGTGCGGAGGAGGAGGAGGAGGAGGAGGAG 560  
QY 121 ProSerSer 123  
|||||

## ORIGIN

## Alignment Scores:

Pred. No.: 4,88e-68 Length: 612  
Score: 639.00 Matches: 123  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x CD676130 (1-612)

QY 1 SerLeupheArgThrValLeuAspSerGlyIleSerGluValArgSerAspAspAsp 20  
Db 6 TCCTCTTCCGACCGTGTGACTCCGACATCTCTGAGTTTCGATCCGACCGGACGACG 65  
QY 21 PheValIlePheLeuAspVallyshisPheSerProGluAspLeuThrVallyValGln 40



```

Db      66  TTGTCATCTTCTCGATGTGAGACATTTCTCCCGAGGAGCCTCACCGTGAGGTGCAG 125
Qy      41  AspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIle 60
Db      126  GACGACTTTGTGGAGATCCACGGAAGCACAAACGAGCGCCAGGACGACCGGCTACATT 185
Qy      61  SerArgGluPheHisArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80
Db      186  TCCCGTAGTTCACACCGCGCTACCGCTCCAAACGTGGACCACTCGCGCCCTCTCT 245
Qy      81  CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100
Db      246  TGCTCCCTGTCTGCGATGGCATGTGACCTTCTGTGCCCCACAGATCCAGACTGGCCTG 305
Qy      101  AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120
Db      306  GATGCCACCCACGCGCGAGGACCATCCCGTGTGCGGGAGGAGAGACCCACCTCGGCT 365
Qy      121  ProSerSer 123
Db      366  CCCTCGTCC 374

```

```

RESULT 27
BF727028
LOCUS      BF727028
DEFINITION by15g05.y1 Human Lens cDNA (Un-normalized, unamplified): BY Homo sapiens cDNA clone by15g05 5', mRNA sequence.
ACCESSION BF727028
VERSION    BF727028.1 GI:12042939
KEYWORDS   EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS      Wistow,G.J., Bernstein,S., Behal,A. and Smith,D.
TITLE        NEIBANK: EST analysis and bioinformatics for ocular genomics
JOURNAL      Invest. Ophthalmol. Vis. Sci. 41 (2000) In press
COMMENT      Contact: Wistow G
              Section on Molecular Structure and Function
              National Eye Institute
              6/331, NIH, Bethesda, MD 20892-2740, USA
              Tel: 301 402 3452
              Fax: 301 496 0078
              Email: graeme@helix.nih.gov
              Plate: 15 row: g column: 05
              Seq primer: M13RP1 reverse primer (ABI).
              Location/Qualifiers
                1. .629
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="by15g05"
                  /tissue_type="Lens"
                  /dev_stage="Adult"
                  /lab_host="EMDH10B"
                  /clone_lib="Human Lens cDNA (Un-normalized, unamplified): BY"
                  /note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses from different adults (both approximately 40 years old) together yielded 20ug of total RNA and 150ng mRNA for cDNA library synthesis. A directionally cloned cDNA library in the pCMVSPORT6 vector was constructed at Life Technologies, essentially following the protocols of the SuperScript Plasmid System full details of which are contained in the manufacturer's instruction manual (http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adaptor [5'-pGACTAGTCTAGATCGGAGCGGCCG(T)15-3']. Not I/blunt end inserts were cloned into the Not I/EcoR V sites in the vector. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

```

## FEATURES

```

source
1. .629
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="taxon:9606"
  /clone="by15g05"
  /tissue_type="Lens"
  /dev_stage="Adult"
  /lab_host="EMDH10B"
  /clone_lib="Human Lens cDNA (Un-normalized, unamplified): BY"
  /note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses from different adults (both approximately 40 years old) together yielded 20ug of total RNA and 150ng mRNA for cDNA library synthesis. A directionally cloned cDNA library in the pCMVSPORT6 vector was constructed at Life Technologies, essentially following the protocols of the SuperScript Plasmid System full details of which are contained in the manufacturer's instruction manual (http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adaptor [5'-pGACTAGTCTAGATCGGAGCGGCCG(T)15-3']. Not I/blunt end inserts were cloned into the Not I/EcoR V sites in the vector. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

```

## ORIGIN

## Alignment Scores:

```

Pred. No.:      5,06e-68      Length:      629
Score:           639.00      Matches:     123
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:     100.00%      Indels:     0
DB:              2           Gaps:         0

```

US-10-657-740-1\_COPY\_51\_173 (1-123) x BF727028 (1-629)

```

Qy      1  SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20
Db      198  TCCCTCTTCCGACCGCTGCTGACTCCGCGCATCTCTGAGGTTGCGATCGACCGGACAAG 257
Qy      21  PheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrValLysValGln 40
Db      258  TTCGTCATCTTCTCGATGTGAGACATTTCTCCCGGAGGACCTCACCGTGNAGGTGCAG 317
Qy      41  AspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIle 60
Db      318  GACGACTTTGTGGAGATCCACGGAAGCACAAACGAGCGCCAGGACGACCGGCTACATT 377
Qy      61  SerArgGluPheHisArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80
Db      378  TCCCGTAGTTCACACCGCGCTACCGCTCCGCTCCAAACGTGGACGACGCGCCCTCTCT 437
Qy      81  CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100
Db      438  TGCTCCCTGTCTGCGATGGCATGTGACCTTCTGTGCCCCACAGATCCAGACTGGCCTG 497
Qy      101  AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120
Db      498  GATGCCACCCACGCGCGAGGACCATCCCGTGTGCGGGAGGAGAGACCCACCTCGGCT 557
Qy      121  ProSerSer 123
Db      558  CCCTCGTCC 566

```

## RESULT 28

```

LOCUS      BM696651
DEFINITION UI-E-DW0-agk-k-24-0-UI.r1 UI-E-DW0 Homo sapiens cDNA clone
ACCESSION  BM696651
VERSION     BM696651.1 GI:19009909
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS      Ronaldo,M.F., Lennon,G. and Soares,M.B.
TITLE        Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL      Genome Res. 6 (9), 791-806 (1996)
MEDLINE     97044477
PUBMED      8889548
COMMENT      Contact: Soares, MB
              Coordinated Laboratory for Computational Genomics
              University of Iowa
              375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
              Tel: 319 335 8250
              Fax: 319 335 9565
              Email: bento-soares@uiowa.edu
              Tissue Procurement: Dr. Gregg Hageman
              cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
              cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
              DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
              Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
              Seq primer: M13 Reverse.

```

FEATURES  
source

Location/Qualifiers  
 1. .631  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-E-DW0-agh-k-24-0-UI"  
 /tissue\_type="lens"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /clone\_lib="UI-E-DW0"  
 /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-DW0 is a cDNA library containing the following tissue(s): lens. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CGATTAGCGA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

## ORIGIN

Alignment Scores:  
 Pred. No.: 5,07e-68 Length: 630  
 Score: 639.00 Matches: 123  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 4 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x BM696651 (1-630)

QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20  
 DB TCCCTCTTCGCGACCGTGGACTCCGGCATCTCTGAGTTCGATCCGCGGACCAAG 263  
 QY 21 PheValIlePheLeuAspValIleHisPheSerProGluAspLeuThrValIleValGln 40  
 DB TTGCTCATCTTCTCGATGTGAAGCACTTCTCCCGGAGGACCTCACCCTGAAGGTGCAG 323  
 QY 41 AspAspPheValGluIleHisGlyIleHisAsnGluArgGlnAspHisGlyTyrIle 60  
 DB GACGACTTTGTGGAGATCCACGGAAGCACACGAGCGCCAGGACGACCGCTACATT 383  
 QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
 DB TCCCTGTAGTTCCACCGCGCTTACCCTGCGCTCCCAACGTCGAGGACCGCTCTCT 443  
 QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100  
 DB TGCTCCCTGTCTGCGGATGGCATGTGACCTTCTGTGCCCCCAAGATCCAGCTGGCGTG 503  
 QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120  
 DB GATGCCACCCAGCCGAGCGAGCCATCCCTGTGTCGGGGAGGAGAGCCACCTCGGCT 563  
 QY 121 ProSerSer 123  
 DB CCGCTGCTCC 572

## RESULT 29

BM706251

## LOCUS

DEFINITION  
 UI-E-DW0-agh-g-05-0-UI.r1 UI-E-DW0 Homo sapiens cDNA clone  
 UI-E-DW0-agh-g-05-0-UI 5', mRNA sequence.

## ACCESSION

BM706251

## VERSION

BM706251.1

## KEYWORDS

EST.

SOURCE  
ORGANISM

Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

## AUTHORS

## TITLE

1 (bases 1 to 631)  
 Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 Normalization and subtraction: two approaches to facilitate gene  
 discovery

## JOURNAL

## MEDLINE

## PUBMED

## COMMENT

Genome Res. 6 (9), 791-806 (1996)  
 97044477  
 8889548  
 Contact: Soares, MB  
 Coordinated Laboratory for Computational Genomics  
 University of Iowa  
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: bento-soares@uiowa.edu  
 Tissue Procurement: Dr. Gregg Hageman  
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research  
 Genetics (www.resgen.com).  
 Seq primer: M13 Reverse.

## FEATURES

## source

Location/Qualifiers  
 1. .631  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-E-DW0-agh-g-05-0-UI"  
 /tissue\_type="lens"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /clone\_lib="UI-E-DW0"  
 /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-DW0 is a cDNA library containing the following tissue(s): lens. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CGATTAGCGA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

## ORIGIN

Alignment Scores:  
 Pred. No.: 5.08e-68 Length: 631  
 Score: 639.00 Matches: 123  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 4 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x BM706251 (1-631)

QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20  
 DB TCCCTCTTCGCGACCGTGGACTCCGGCATCTCTGAGTTCGATCCGCGGACCAAG 109  
 QY 21 PheValIlePheLeuAspValIleHisPheSerProGluAspLeuThrValIleValGln 40  
 DB TTGCTCATCTTCTCGATGTGAAGCACTTCTCCCGGAGGACCTCACCCTGAAGGTGCAG 169  
 QY 41 AspAspPheValGluIleHisGlyIleHisAsnGluArgGlnAspHisGlyTyrIle 60  
 DB GACGACTTTGTGGAGATCCACGGAAGCACACGAGCGCCAGGACCGCTACATT 229



```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-DW0-agh-g-03-0-UI"
/tissue_type="lens"
/dev_stage="adult"
/lab_host="DHI0B (Life Technologies) (T1 phage resistant)"
/notes="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-E-DW0 is a cDNA library containing the following tissue(s): lens. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CGATTAGCGA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."
```

## ORIGIN

## Alignment Scores:

```

Pred. No.: 5,12e-68 Length: 635
Score: 639.00 Matches: 123
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0
```

US-10-657-740-1\_COPY\_51\_173 (1-123) x BM697066 (1-635)

```

QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20
Db 148 TCCTCTTCGCGACCGTGTGAGCTCCGGATCTCTGAGGTTCGATCCGACGGGACAA 207
QY 21 PheValIlePheLeuAspValIleHisGlyLysHisAsnGluArgGlnAspHisGlyTyrIle 60
Db 208 TTGCTCATCTTCTCGATGTGAAGCACTTCTCCCGAGAGACTTACCGTGAAGGTGAG 267
QY 41 AspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspHisGlyTyrIle 60
Db 268 GAGCACTTTGTGAGATCCACGGAAGACACAGACGCGGACGACGACGCGCTACATT 327
QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80
Db 328 TCCCGTGAGTTCCACCGCGCTTACCGCTGCCGTCCAACTGGACGACGTGGCGCTCTCT 387
QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100
Db 388 TGCTCCCTCTCTGCGCATGGCATGCTGACCTTCTGTGCCCCAAGATCCAGCTGGCTG 447
QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120
Db 448 GATGCCACCCAGCCGAGCGGATCCCTCCCGTGTGCGGGAGGAGAGCCACCTCGGCT 507
QY 121 ProSerSer 123
Db 508 CCCTCGTCC 516
```

## RESULT 32

## BM706270

## LOCUS

DEFINITION UI-E-DW0-agh-k-09-0-UI.r1 UI-E-DW0 Homo sapiens linear EST 28-FEB-2002

ACCESSION UI-E-DW0-agh-k-09-0-UI 5', mRNA sequence.

VERSION BM706270

KEYWORDS BM706270.1 GI:19019528

SOURCE EST.

ORGANISM Homo sapiens (human)

1 SerArgGluPheHisArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

## AUTHORS

## TITLE

1 (bases 1 to 659)

Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

8889548

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

Tissue Procurement: Dr. Gregg Hageman

cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research

Genetics (www.resgen.com).

Seq primer: M13 Reverse.

Location/Qualifiers

1..659

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="UI-E-DW0-agh-k-09-0-UI"

/tissue\_type="lens"

/dev\_stage="adult"

/lab\_host="DHI0B (Life Technologies) (T1 phage resistant)"

/clone\_lib="UI-E-DW0"

/notes="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a

modified polylinker; Site\_1: EcoR I; Site\_2: Not I;

UI-E-DW0 is a cDNA library containing the following

tissue(s): lens. The library was constructed according to

Bonaldo, Lennon and Soares, Genome Research, 6:791-806,

1996. First strand cDNA synthesis was primed with an

oligo-dT primer containing a Not I site. Double stranded

cDNA was ligated to an EcoR I adaptor, digested with Not

I, and cloned directionally into pT7T3-Pac vector. The

oligonucleotide used to prime the synthesis of

first-strand cDNA contains a library tag sequence that is

located between the Not I site and the (dT)18 tail. The

sequence tag for this library is CGATTAGCGA. This library

was created for the program, Gene Discovery in the Visual

System, supported by National Eye Institute (NEI)."

## ORIGIN

## Alignment Scores:

```

Pred. No.: 5.38e-68 Length: 659
Score: 639.00 Matches: 123
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0
```

US-10-657-740-1\_COPY\_51\_173 (1-123) x BM706270 (1-659)

```

QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20
Db 207 TCCTCTTCGCGACCGTGTGAGCTCCGGATCTCTGAGGTTCGATCCGACGGGACAA 266
QY 21 PheValIlePheLeuAspValIleHisGlyLysHisAsnGluArgGlnAspHisGlyTyrIle 60
Db 267 TTGCTCATCTTCTCGATGTGAAGCACTTCTCCCGAGAGACTTACCGTGAAGGTGAG 326
QY 41 AspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspHisGlyTyrIle 60
Db 327 GAGCACTTTGTGAGATCCACGGAAGACACAGACGCGGACGACGCTACATT 386
QY 61 SerArgGluPheHisArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80
```

```

Db      387 TCCCGTGGTTCACCGCGCTACCGCTCGCCGTCGACGTCGACGTCGCGCCCTCTCT 446
Qy      81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysileGlnThrGlyLeu 100
Db      447 TGCTCCCTGCTGCGGATGGCATGCTGACCTTCTGTGGCCCAAGATCCAGACTGCGCTG 506
Qy      101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120
Db      507 GATGCCACCCACCGCGAGCGACCATCCCGTGTGCGGGAGGAGGAGCCACCTCGGCT 566
Qy      121 ProSerSer 123
Db      567 CCCTCGTCC 575

RESULT 33
LOCUS   BM697160
DEFINITION UI-E-DW0-aggm-i-06-0-UI-r1 UI-E-DW0 Homo sapiens cDNA clone
ACCESSION BM697160
VERSION    1
KEYWORDS   Homo sapiens (human)
SOURCE     Homo sapiens
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 661)
AUTHORS    Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE      Normalization and subtraction: two approaches to facilitate gene
JOURNAL    Genome Res. 6 (9), 791-806 (1996)
MEDLINE    9704477
PUBMED     8889548
COMMENT    Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@iowa.edu
            Tissue Procurement: Dr. Gregg Hageman
            cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            Seq primer: M13 Reverse.
            Location/Qualifiers
            1..661
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /clone="UI-E-DW0-aggm-i-06-0-UI"
               /tissue_type="lens"
               /dev_stage="adult"
               /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
               /clone_lib="UI-E-DW0"
               /note="organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
            modified polylinker; Site_1: EcoR I; Site_2: Not I;
            UI-E-DW0 is a cDNA library containing the following
            tissue(s): lens. The library was constructed according to
            Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
            1996. First strand cDNA synthesis was primed with an
            oligo-dT primer containing a Not I site. Double stranded
            cDNA was ligated to an EcoR I adaptor, digested with Not
            I, and cloned directionally into pT7T3-Pac vector. The
            oligonucleotide used to prime the synthesis of
            first-strand cDNA contains a library tag sequence that is
            located between the Not I site and the (dT)18 tail. The
            sequence tag for this library is CGATTAGCGA. This library
            was created for the program, Gene Discovery in the Visual
            System, supported by National Eye Institute (NEI)."
FEATURES
source

```

## ORIGIN

## Alignment Scores:

Pred. No.: 5,4e-68 Length: 661  
 Score: 639.00 Matches: 123  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 4 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x BM697160 (1-661)

Qy 1 SerLeuPheArgThrValLeuAspSerGlylleSerGluValArgSerAspArgAspLys 20  
 Db 144 TCCCTCTTCCCGCACCGCTGCTGACTCCCGCATCTCTGAGGTTTCGATCCGACCGGGCAAG 203  
 Qy 21 PheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrValLysValGln 40  
 Db 204 TTCGTCATCTTCTCGATGTGAGCACCTTCTCCCGGAGGAGCTCACCGTGAAGGTGCAG 263  
 Qy 41 AspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIle 60  
 Db 264 GACGACTTTGTGGAGATCCACGGAAGCACACAGAGCGCCAGACGACCGCTACATT 323  
 Qy 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
 Db 324 TCCCGTGTAGTTCCACCGCGCTACCGCTGCGCGTCCAACTGAGGACGACGCTCGCTCTCT 383  
 Qy 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysileGlnThrGlyLeu 100  
 Db 384 TGCTCCCTGCTGCGGATGGCATGCTGACCTTCTGTGGCCCCAGATCCAGACTCGCCTG 443  
 Qy 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120  
 Db 444 GATGCCACCCACCGCGAGCGACCATCCCGTGTGCGGAGGAGGAGCCACCTCGGCT 503  
 Qy 121 ProSerSer 123  
 Db 504 CCCTCGTCC 512

## RESULT 34

## LOCUS

## DEFINITION

BM697160 Soares retina N2b4HR Homo sapiens cDNA clone

IMAGE:220285, mRNA sequence.

## ACCESSION

BM697160

## VERSION

1

## KEYWORDS

EST.

## SOURCE

Homo sapiens (human)

## ORGANISM

Homo sapiens

## REFERENCE

1 (bases 1 to 661)

## AUTHORS

Radelof,U., Schneider,D. and Korn,B.

## TITLE

Human Unigeneset - RZPD3

## JOURNAL

Unpublished (2003)

## COMMENT

Contact: Ina Rolfs

RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH

Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany

RZPD: IMAGp998PI4436.

RZPDLIB; I.M.A.G.E. cDNA Clone Collection;

Human Unigeneset - RZPD3 (RZPDLIB No.972)

http://www.rzpd.de/CloneCards/cgi-

bin/showLib.pl.cgi/response?libNo=972 Contact: Ina Rolfs

RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH

Heubnerweg 6, D-14059 Berlin, Germany

Tel: +49 30 32639 101

Fax: +49 30 32639 111

www.rzpd.de

This clone is available royalty-free from RZPD;

contact RZPD (clone@rzpd.de) for further information. Seq primer:

M13r, Primer sequence: TTTACACAGAAACAGCTATGAC.

## FEATURES

## source

FEATURES  
source

Location/Qualifiers  
1. .681  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE998P14436 ; IMAGE:220285"  
/sex="male"  
/tissue\_type="retina"  
/dev\_stage="55 year old"  
/lab\_host="DH10B (ampicillin resistant)"  
/clone\_lib="Soares retina N2b4HR"  
/notes="Organ: eye; Vector: pTV73D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCGAAGTGGAGCGCGCCGCTTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pTV73 vector (Pharmacia). The retinas were obtained from a 55 year old Caucasian and total cellular poly(A)+ RNA was extracted 6 hrs after their removal. The retina RNA was kindly provided by Roderick R. McInnes M.D. Ph.D. from the University of Toronto. Library constructed by Bento Soares and M.Fatima Donaldso. "

## ORIGIN

## Alignment Scores:

Pred. No.: 5.61e-68 Length: 681  
Score: 639.00 Matches: 123  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 5 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x BX118596 (1-681)

QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20  
Db 243 TCCCTCTTCGCGACCGTGGACTCCGGCATCTCTGAGGTTCTGATCCGACCGGACAA 302  
QY 21 PheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrValLysValGln 40  
Db 303 TTGCTCATCTTCTCGATGTGAAGCACTTCTCCCGAGGACCTCACCCTGAAGGTGCAG 362  
QY 41 AspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrlle 60  
Db 363 GACGACTTTGTGGAGATCCACGGAAGACACACGAGCGCCAGACGACGCTACATT 422  
QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
Db 423 TCCCGTGAGTTCCACCGCGCTTACCGCTCGCTCCACGTTGGACCGACCTCTCTCT 482  
QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100  
Db 483 TGCTCCCTCTCTCGCGATGGCATGTGACCTTCTGTGGCCCAAGATCCAGACTGGCGTG 542  
QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120  
Db 543 GATGCCACCGACCGGAGCGACCATCCCGTGTCGCGGGAGGAGACCCACCTCGGCT 602  
QY 121 ProSerSer 123  
Db 603 CCTCTGCTCC 611

## RESULT 35

BF727324

LOCUS

DEFINITION

by19h12.y1 Human Lens cDNA (Un-normalized, EST 05-JAN-2001

sapiens cDNA clone by19h12 5', mRNA sequence. BY Homo

ACCESSION

BF727324

VERSION

BF727324.1 GI:12043235

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 698)

Wistow,G.J., Bernstein,S., Behal,A. and Smith,D.

NEIBANK: EST analysis and bioinformatics for ocular genomics

Invest. Ophthalmol. Vis. Sci. 41 (2000) In press

Contact: Wistow G

Section on Molecular Structure and Function

National Eye Institute

6/331, NIH, Bethesda, MD 20892-2740, USA

Tel: 301 402 3452

Fax: 301 496 0078

Email: graeme@helix.nih.gov

Plate: 19 row: h column: 12

Seq primer: M13RP1 reverse primer (ABI).

Location/Qualifiers

1. .698

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="by19h12"

/tissue\_type="Lens"

/dev\_stage="Adult"

/lab\_host="EMDH10B"

/clone\_lib="Human Lens cDNA (Un-normalized, unamplified):

By"

/notes="Organ: Eye; Vector: pCMVSPORT6; Two human lenses

from different adults (both approximately 40 years old)

together yielded 20ug of total RNA and 150ng mRNA for cDNA

library synthesis. A directionally cloned cDNA library in

the pCMVSPORT6 vector was constructed at Life

Technologies, essentially following the protocols of the

SuperScript Plasmid System full details of which are

contained in the manufacturer's instruction manual

(http://www.lifetech.com/). First strand synthesis was

carried out using a Not I primer-adaptor

[5'-pOACTAGTCTTAGATCGGAGCGCGCC(T)15-3']. Not I/blunt

end inserts were cloned into the Not I/EcoR V sites in the

vector. EST analysis was performed on the unamplified

library at the NIH Intramural Sequencing Center (NISC)."

## ORIGIN

## Alignment Scores:

Pred. No.: 5.8e-68 Length: 698  
Score: 639.00 Matches: 123  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 2 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x BF727324 (1-698)

QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20  
Db 205 TCCCTCTTCGCGACCGTGGACTCCGGCATCTCTGAGGTTCTGATCCGACCGGACAA 264  
QY 21 PheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrValLysValGln 40  
Db 265 TTGCTCATCTTCTCGATGTGAAGCACTTCTCCCGGAGGACCTCACCCTGAAGGTGCAG 324  
QY 41 AspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrlle 60  
Db 325 GACGACTTTGTGGAGATCCACGGAAGACACACGAGCGCCAGGACGACCATCATTT 384  
QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
Db 385 TCCCGTGAGTTCCACCGCGCTTACCGCTCGCTCCACGTCGACGACGCGCTCTCT 444  
QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100  
Db 445 TGCTCCCTCTCTCGCGATGGCATGCTTGACTTCTGTGGCCCAAGATCCAGATCGCCTG 504



```

/dev_stage="fetal"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-E00"
/notes="Organ: eye; Vector: pTT73-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-E-E00 is a cDNA library containing the following
tissue(s): fetal eye. The library was constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pTT73-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tag for this library is CCGGATACC. This library
was created for the program, Gene Discovery in the Visual
System, supported by National Eye Institute (NEI)."
```

## ORIGIN

```

Alignment Scores:
Pred. No.: 1,35e-67 Length: 559
Score: 635.00 Matches: 122
Percent Similarity: 99.19% Conservative: 0
Best Local Similarity: 99.19% Mismatches: 1
Query Match: 99.37% Indels: 0
DB: 4 Gaps: 0
```

```
US-10-657-740-1_COPY_51_173 (1-123) x BM722779 (1-559)
```

```

QY 1 SerLeupheArghThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20
DB 186 TCCCTCTTCCGACCGTCTCGGATCTCTGAGTTCTGATCCGCGGCAAG 245
QY 21 PheValIlePheLeuAspValIleHisPheSerProGluAspLeuThrValIleValGln 40
DB 246 TTGTCATCTTCTCGATGTGAACACTTCTCCCGGAGGACCTCACCGTGAAGTGCAG 305
QY 41 AspAspPheValGluIleHisGlyIleHisAsnGluArgGlnAspAspHisGlyTyrlle 60
DB 306 GACGACTTTGTGGAGATCCACGGAAGCAACAGCGCCAGGACGACCATTCGAGTGCAG 365
QY 61 SerArgGluPheHisArgArgTyArgLeuProSerAsnValAspGlnSerAlaLeuSer 80
DB 366 TCCCGTGAGTTCCACCGCGCTACCGCTCGCGTCCACGTCGACGACCTCTCTCT 425
QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProIleGlnThrGlyLeu 100
DB 426 TGCTCCCTGTCCTCGCATGGCATCTCTGTGGCCCAAGATCCAGCTGGCTG 485
QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120
DB 486 GATGCCACCCAGCGGAGGACCATCCCGTGTCCGCGGAGGAGAGCCACCTCGGNT 545
QY 121 ProSerSer 123
DB 546 CCTCGTCC 554
```

## RESULT 38

```

BF726358
LOCUS BF726358
DEFINITION By05d12.y1 Human Lens cDNA (Un-normalized, unamplified); BY Homo
sapiens cDNA clone by05d12 5', mRNA sequence.
ACCESSION BF726358
VERSION BF726358.1 GI:12042269
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
```

```

REFERENCE 1 (bases 1 to 532)
AUTHORS Wistow,G.J., Bernstein,S., Behal,A. and Smith,D.
TITLE NEIBANK: EST analysis and bioinformatics for ocular genomics
```

## JOURNAL COMMENT

Invest. Ophthalmol. Vis. Sci. 41 (2000) In press  
 Contact: Wistow G  
 Section on Molecular Structure and Function  
 National Eye Institute  
 6/331, NIH, Bethesda, MD 20892-2740, USA  
 Tel: 301 402 3452  
 Fax: 301 496 0078  
 Email: grameehelix.nih.gov  
 Plate: 95 row: d column: 12  
 Seq primer: M13RP1 reverse primer (ABI).

## FEATURES

## source

```

1..532
Location/Qualifiers
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="taxon:9606"
  /clone="by05d12"
  /tissue_type="Lens"
  /dev_stage="Adult"
  /lab_host="EMDH10B"
  /clone_lib="Human Lens cDNA (Un-normalized, unamplified)"
  By"
```

/note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses from different adults (both approximately 40 years old) together yielded 20ug of total RNA and 150ng mRNA for cDNA library synthesis. A directionally cloned cDNA library in the pCMVSPORT6 vector was constructed at Life Technologies, essentially following the protocols of the SuperScript Plasmid System full details of which are contained in the manufacturer's instruction manual (http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adaptor [5'-pGACAGTCTAGATCGGAGCGCGCC(T)15-3']. Not I/blunt end inserts were cloned into the Not I/EcoR V sites in the vector. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

## ORIGIN

```

Alignment Scores:
Pred. No.: 2,22e-67 Length: 532
Score: 633.00 Matches: 122
Percent Similarity: 99.19% Conservative: 0
Best Local Similarity: 99.19% Mismatches: 1
Query Match: 99.06% Indels: 0
DB: 2 Gaps: 0
```

```
US-10-657-740-1_COPY_51_173 (1-123) x BF726358 (1-532)
```

```

QY 1 SerLeupheArghThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20
DB 152 TCCCTCTTCCGACCGTCTCGGATCTCTGAGTTCTGATCCGCGGCAAG 211
QY 21 PheValIlePheLeuAspValIleHisPheSerProGluAspLeuThrValIleValGln 40
DB 212 TTGTCATCTTCTCGATGTGAACACTTCTCCCGGAGGACCTCACCGTGAAGTGCAG 271
QY 41 AspAspPheValGluIleHisGlyIleHisAsnGluArgGlnAspAspHisGlyTyrlle 60
DB 272 GACGACTTTGTGGAGATCCACGGAAGCAACAGCGCCAGGACGACCATTCAT 331
QY 61 SerArgGluPheHisArgArgTyArgLeuProSerAsnValAspGlnSerAlaLeuSer 80
DB 332 TCCCGTGAGTTCCACCGCGCTACCGCTCGCGTCCACGTCGACCATTCCTCTCT 391
QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProIleGlnThrGlyLeu 100
DB 392 TGCTCCCTGTCCTCGCATGGCATCTGACCTTCTGTGGCCCAAGATCCAGCTG 451
QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120
DB 452 GATGCCACCCAGCGGAGGACCATCCCGTGTCCGCGGAGGAGAGCCACCTCGGCT 511
QY 121 ProSerSer 123
DB 511
```



Db 512 CCCTCGTCC 520

RESULT 39  
AY419530  
LOCUS  
DEFINITION Pan troglodytes CRYAA gene, VIRTUAL TRANSCRIPT, partial sequence,  
genomic survey sequence.  
ACCESSION AY419530  
VERSION AY419530.1 GI:39775487  
KEYWORDS GSS.  
SOURCE Pan troglodytes (chimpanzee)  
ORGANISM Pan troglodytes  
REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
1 (bases 1 to 522)  
Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarawal, A.,  
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,  
Ferrera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,  
Adams, M.D. and Cargill, M.  
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous  
gene trios  
JOURNAL Science 302 (5652), 1960-1963 (2003)  
PUBMED 14671302  
REFERENCE  
AUTHORS 2 (bases 1 to 522)  
Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarawal, A.,  
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,  
Ferrera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,  
Adams, M.D. and Cargill, M.  
TITLE Direct Submission  
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
Rockville, MD 20850, USA  
COMMENT These sequences were made by sequencing genomic exons and ordering  
them based on alignment.  
FEATURES  
source  
1..522  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
<1..>522  
/gene="CRYAA"  
/locus\_tag="HCW6921"

gene

ORIGIN

Alignment Scores:  
Pred. No.: 2.88e-67 Length: 522  
Score: 632.00 Matches: 122  
Percent Similarity: 99.19% Conservative: 0  
Best Local Similarity: 99.19% Mismatches: 1  
Query Match: 98.90% Indels: 0  
DB: 9 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x AY419530 (1-522)

Qy 1 SerLeuPheArgThrValLeuAspSerGlyLeuValArgSerAspArgAspLys 20  
Db 151 TCCCTCTTCCGACCGCTGCGACTCCGGCATCTCTGAGGTTCCGATCCGCGGACAAG 210  
Qy 21 PheValIlePheLeuAspValIleHisPheSerProGluAspLeuThrValIleValGln 40  
Db 211 TTGTCATCTTCTTCGATGTGAAGACATCTCCCGGAGGACCTCACCCTGAAGGTGCAG 270  
Qy 41 AspAspPheValGluIleHisGlyIleHisAsnGluArgGlnAspHisGlyTyrIle 60  
Db 271 GACGACTTTGTGGAGATCCACGGAAGCACACGAGCGCCAGNNNGACCGCTACATT 330  
Qy 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
Db 331 TCCGTGAGTTCACCGCGCTACCGCTCCACGTCGACGACGACGCTCTCTCT 390  
Qy 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProIleGlnThrGlyLeu 100  
Db 391 TGCTCCCTGTCCGCGGATGGCATGCTGACCTTCTGTGTGCCCCCAAGATCCAGATGGCCTG 450

Qy 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120  
Db 451 GATGCCACCCAGCCGAGCCATCCCGTGTCCGGGAGGAGAGCCACCTCGGCT 510  
Qy 121 ProSerSer 123  
Db 511 CCCTCGTCC 519  
RESULT 40  
BM696477  
LOCUS  
DEFINITION UI-E-DW0-agj-1-20-0-UI.r1 UI-E-DW0 Homo sapiens cDNA clone  
UI-E-DW0-agj-1-20-0-UI 5', mRNA sequence.  
ACCESSION BM696477  
VERSION BM696477.1 GI:19009735  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 519)  
Bonaldo, M.F., Lennon, G. and Soares, M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene  
discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 8889548  
PUBMED 97044477  
COMMENT Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@uiowa.edu  
Tissue Procurement: Dr. Gregg Hageman  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics (www.resgen.com).  
Seq primer: M13 Reverse.  
FEATURES  
source  
Location/Qualifiers  
1..519  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-E-DW0-agj-1-20-0-UI"  
/tissue\_type="lens"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone\_lib="UI-E-DW0"  
/note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a  
modified polylinker; Site\_1: EcoR I; Site\_2: Not I;  
UI-E-DW0 is a cDNA library containing the following  
tissues(s): lens. The library was constructed according to  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. First strand cDNA synthesis was primed with an  
oligo-dT primer containing a Not I site. Double stranded  
cDNA was ligated to an EcoR I adaptor, digested with Not  
I, and cloned directionally into pT7T3-Pac vector. The  
oligonucleotide used to prime the synthesis of  
first-strand cDNA contains a library tag sequence that is  
located between the Not I site and the (dT)18 tail. The  
sequence tag for this library is CGATTAGCGA. This library  
was created for the program, Gene Discovery in the Visual  
System, supported by National Eye Institute (NEI)."

ORIGIN

Alignment Scores:  
Pred. No.: 3.79e-67 Length: 519  
Score: 631.00 Matches: 122  
Percent Similarity: 99.19% Conservative: 0  
Best Local Similarity: 99.19% Mismatches: 1

```

Query Match: 98.75% Indels: 0
DB: 4 Gaps: 0
US-10-657-740-1_COPY_51_173 (1-123) x BM696477 (1-519)

QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspIlys 20
DB 135 TCCCTCTTCCGACCGTCTGGACTCCGGCATCTCTGAGGTTCCATCCGACCGGCAAG 194
QY 21 PheValIlePheLeuAspValIyHisPheSerProGluAspLeuThrValIyValGln 40
DB 195 TTCGTCATCTTCTCCGATGTGAAGCACCTCTCCCGGAGGACCTCACCGTGAAGTGCAG 254
QY 41 AspAspPheValGluIleHisGlyIyHisAsnGluArgGlnAspAspHisGlyIyTyrIle 60
DB 255 GACGACTTTGTGGAGATCCACGGAAGACCAACGAGCGCCAGGACGACCGGCTACATT 314
QY 61 SerArgGluPheHisArgTyArgIleProSerAsnValAspGlnSerAlaLeuSer 80
DB 315 TCCCGTGAGTTCCACCGCGCTACCGCTGCGTCCCAACGTGGACCACTCGGCCCTCTCT 374
QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProIySileGlnThrGlyLeu 100
DB 375 TGCTCCCTGCTGCGATGGCATGCTGACCTTCTGTGCCCCCAAGATCCAGACTGGCCTG 434
QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluIyLysProThrSerAla 120
DB 435 GATGCCACCCACCGCGAGGAGCCATCCCGTGTGCGGGGAGGAGAAAGCCACTCGGCT 494
QY 121 ProSerSer 123
DB 495 CCCTCGTCC 503

RESULT 41
BF726253
LOCUS
DEFINITION
by03h05.y1 Human Lens cDNA (Un-normalized, unamplified): BY Homo
ACCESSION
BF726253
VERSION
BF726253.1 GI:12042164
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 577)
Wistow,G.J., Bernstein,S., Behal,A. and Smith,D.
NEIBANK: EST analysis and bioinformatics for ocular genomics
Invest. Ophthalmol. Vis. Sci. 41 (2000) In press
Contact: Wistow G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graeme@helix.nih.gov
Plate: 03 row: h column: 05
Seq primer: M13Rpl reverse primer (ABI).
FEATURES
Location/Qualifiers
1..577
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="by03h05"
/tissue_type="Lens"
/dev_stage="Adult"
/lab_host="EMDH10B"
/clone_lib="Human Lens cDNA (Un-normalized, unamplified):
By"
/note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses
from different adults (both approximately 40 years old)
together yielded 20ug of total RNA and 150ng mRNA for cDNA
library synthesis. A directionally cloned cDNA library in

```

the pCMVSPORT6 vector was constructed at Life Technologies, essentially following the protocols of the SuperScript Plasmid System full details of which are contained in the manufacturer's instruction manual (<http://www.lifetech.com/>). First strand synthesis was carried out using a Not I primer-adaptor [5'-TGACTAGTTCTAGATCGGAGCGCCGCTT]5'-3']. Not I/blunt end inserts were cloned into the Not I/EcoR V sites in the vector. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC).

## ORIGIN

```

Alignment Scores:
Pred. No.: 5,78e-67 Length: 577
Score: 630.00 Matches: 122
Percent Similarity: 99.1% Conservative: 0
Best Local Similarity: 99.1% Mismatches: 1
Query Match: 98.59% Indels: 0
DB: 2 Gaps: 0
US-10-657-740-1_COPY_51_173 (1-123) x BF726253 (1-577)
QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspIlys 20
DB 198 TCCCTCTTCCGACCGTCTGGACTCCGGCATCTCTGAGGTTCCATCCGACCGGCAAG 257
QY 21 PheValIlePheLeuAspValIyHisPheSerProGluAspLeuThrValIyValGln 40
DB 258 TTCGTCATCTTCTCCGATGTGAAGCACCTTCTCCCGGAGGACCTCACCGTGAAGTGCAG 317
QY 41 AspAspPheValGluIleHisGlyIyHisAsnGluArgGlnAspAspHisGlyIyTyrIle 60
DB 318 GACGACTTTGTGGAGATCCACGGAAGACCAACGAGCGCCAGGACCGGCTACATT 377
QY 61 SerArgGluPheHisArgTyArgLeuProSerAsnValAspGlnSerAlaLeuSer 80
DB 378 TCCCGTGAGTTCCACCGCGCTACCGCTGCGTCCCAACGTGGACCACTCGGCCCTCTCT 437
QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProIySileGlnThrGlyLeu 100
DB 438 TGCTCCCTGCTGCGATGGCATGCTGACCTTCTGTGCCCCCAAGATCCAGACTGGCCTG 497
QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluIyLysProThrSerAla 120
DB 498 GATGCCACCCACCGCGAGGAGCCATCCCGTGTGCGGGGAGGAGAAAGCCACTCGGCT 557
QY 121 ProSerSer 123
DB 558 CCCTCGTCC 566

```

```

BF726422 587 bp mRNA linear EST 05-JAN-2001
by06d05.y1 Human Lens cDNA (Un-normalized, unamplified): BY Homo
BF726422
BF726422.1 GI:12042333
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 587)
Wistow,G.J., Bernstein,S., Behal,A. and Smith,D.
NEIBANK: EST analysis and bioinformatics for ocular genomics
Invest. Ophthalmol. Vis. Sci. 41 (2000) In press
Contact: Wistow G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graeme@helix.nih.gov

```

Plate: 06 row: d column: 05  
Seq primer: M13RPI reverse primer (ABI).  
Location/Qualifiers  
1. .587

# FEATURES

source  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="by06d05"  
/tissue\_type="Lens"  
/dev\_stage="Adult"  
/lab\_host="EMDH10B"  
/clone\_lib="Human Lens cDNA (Un-normalized, unamplified):  
By"  
/note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses  
from different adults (both approximately 40 years old)  
together yielded 20ug of total RNA and 150ng mRNA for cDNA  
library synthesis. A directionally cloned cDNA library in  
the pCMVSPORT6 vector was constructed at Life  
Technologies, essentially following the protocols of the  
SuperScript Plasmid System full details of which are  
contained in the manufacturer's instruction manual  
(http://www.lifetech.com/). First strand synthesis was  
carried out using a Not I primer-adaptor  
[5'-pGACTAGTCTAGATCGGCGGCCGCC(T)15-3']. Not I/blunt  
end inserts were cloned into the Not I/EcoR V sites in the  
vector. EST analysis was performed on the unamplified  
library at the NIH Intramural Sequencing Center (NISC)."

# ORIGIN

Alignment Scores:  
Pred. No.: 5,91e-67 Length: 587  
Score: 630.00 Matches: 122  
Percent Similarity: 99.19% Conservative: 0  
Best Local Similarity: 99.19% Mismatches: 1  
Query Match: 98.59% Indels: 0  
DB: 2 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x BF726422 (1-587)

Qy 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20  
Db 158 TCCCTCTTCGACACGCTGCTGAGCTCCGCGCATCTCTGAGGTTGATCCGCGGACGAC 217  
Qy 21 PheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrValLysValGln 40  
Db 218 TTCGTCTATCTCTCGATGTGAAGCATTCTCCCGGAGGACCTCACCGTGAAGGTGCGAG 277  
Qy 41 AspAspPheValGluIleHisGlyIleHisGluArgGlnAspAspHisGlyTyrIle 60  
Db 278 GACGACTTTGTGGAGATCCAGGAAGCACACACGAGCGCCAGGACGACCGGCTACATTT 337  
Qy 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
Db 338 TCCCGTGAGTTCACCGCGCTACCGCTCGCTCCGCTCAACGTGGACGATCGGCGCTCTCT 397  
Qy 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100  
Db 398 TGCTCCCTGTCTGCCGATGGATGCTGACCTTCTGTGCCCCCAAGATCCAGACTGGCCTG 457  
Qy 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120  
Db 458 GATGCCACCCAGCCGAGCGGCCATCCNCGTGTGCGGGAGGAGAGCCACCTCGGCT 517  
Qy 121 ProSerSer 123  
Db 518 CCCTGTCC 526

# RESULT 43

BF726330  
LOCUS BF726330 597 bp mRNA linear EST 05-JAN-2001  
DEFINITION by05b01.y1 Human Lens cDNA (Un-normalized, unamplified): BY Homo  
sapiens cDNA clone by05b01 5', mRNA sequence.  
ACCESSION BF726330

VERSION BF726330.1 GI:12042241  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 597)  
AUTHORS Wistow,G.J., Bernstein,S., Behal,A. and Smith,D.  
TITLE NEIBANK: EST analysis and bioinformatics for ocular genomics  
JOURNAL Invest. Ophthalmol. Vis. Sci. 41 (2000) In press  
COMMENT Contact: Wistow G  
Section on Molecular Structure and Function  
National Eye Institute  
6/331, NIH, Bethesda, MD 20892-2740, USA  
Tel: 301 402 3452  
Fax: 301 496 0078  
Email: graeme@helix.nih.gov  
Plate: 05 row: b column: 01  
Seq primer: M13RPI reverse primer (ABI).  
Location/Qualifiers  
1. .597

# FEATURES

source  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="by05b01"  
/tissue\_type="Lens"  
/dev\_stage="Adult"  
/lab\_host="EMDH10B"  
/clone\_lib="Human Lens cDNA (Un-normalized, unamplified):  
By"  
/note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses  
from different adults (both approximately 40 years old)  
together yielded 20ug of total RNA and 150ng mRNA for cDNA  
library synthesis. A directionally cloned cDNA library in  
the pCMVSPORT6 vector was constructed at Life  
Technologies, essentially following the protocols of the  
SuperScript Plasmid System full details of which are  
contained in the manufacturer's instruction manual  
(http://www.lifetech.com/). First strand synthesis was  
carried out using a Not I primer-adaptor  
[5'-pGACTAGTCTAGATCGGCGGCCGCC(T)15-3']. Not I/blunt  
end inserts were cloned into the Not I/EcoR V sites in the  
vector. EST analysis was performed on the unamplified  
library at the NIH Intramural Sequencing Center (NISC)."

# ORIGIN

Alignment Scores:  
Pred. No.: 8,02e-67 Length: 597  
Score: 629.00 Matches: 121  
Percent Similarity: 98.37% Conservative: 0  
Best Local Similarity: 98.37% Mismatches: 2  
Query Match: 98.44% Indels: 0  
DB: 2 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x BF726330 (1-597)

Qy 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20  
Db 152 TCCCTCTTCGACACGCTGCTGAGCTCCGCGCATCTCTGAGGTTGATCCGCGGACGAC 211  
Qy 21 PheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrValLysValGln 40  
Db 212 TTCGTCTATCTCTCGATGTGAAGCATTCTCCCGGAGGACCTTCACCGTGAAGGTGCGAG 271  
Qy 41 AspAspPheValGluIleHisGlyIleHisGluArgGlnAspAspHisGlyTyrIle 60  
Db 272 GACGACTTTGTGGAGATCCAGGAAGCACACACGAGCGCCAGGACGACCGGCTACATTT 331  
Qy 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
Db 332 TCCCGTGAGTTCACCGCGCTACCGCTCGCTCCAGCTGACGAGGACGAGTGGCCCTCTCT 391  
Qy 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 100



/dev stage="adult"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /clone\_lib="UI-E-DW0"  
 /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-DW0 is a cDNA library containing the following tissue(s): lens. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CGATTAGCGA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

## ORIGIN

Alignment Scores:  
 Pred. No.: 1,81e-66 Length: 580  
 Score: 626.00 Matches: 120  
 Percent Similarity: 99.17% Conservativeness: 0  
 Best Local Similarity: 99.17% Mismatches: 1  
 Query Match: 97.97% Indels: 0  
 DB: 4 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x BM706139 (1-580)

QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20  
 DB 206 TCCCTCTTCCGACCGTGTGACCTCCGACATCTCTGAGGTTCGATCGACCGGACAAG 265  
 QY 21 PheValIlePheLeuAspVallyshisPheSerProGluAspLeuThrVallyshisValGln 40  
 DB 266 TTGCTCATCTCTCTGATGTGAGACATCTCTCCCGAGGAGCTCCTGAGTGGAGTGACG 325  
 QY 41 AspAspPheValGluIleHisGlylyshisAsnGluArgGlnAspAspHisGlyTyrIle 60  
 DB 326 GACGACTTTGTGGAGATCCAGGAAGCACACAGAGCGCCAGGACGACCGCTTACATT 385  
 QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
 DB 386 TCCCGTGAGTTCCACCGCGCTTACCGCTCCGCTCAACAGTGGACGACGCTGCTCTCT 445  
 QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLyshisGlnThrGlyLeu 100  
 DB 446 TGCTCNCCTGTGCGGATGGCATGTGACCTTCTGTGGCCCCAAGATCCAGACTGGGCTG 505  
 QY 101 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 120  
 DB 506 GATGCCACCCAGCGGAGGACCATCCCGTGTCCGGGAGGAGAGCCCACTCGCT 565  
 QY 121 Pro 121  
 DB 566 CCC 568

## RESULT 46

BM722650  
 LOCUS BM722650 586 bp mRNA linear EST 01-MAR-2002  
 DEFINITION UI-E-E00-ahy-a-10-0-UI.r1 UI-E-E00 Homo sapiens cDNA clone  
 UI-E-E00-ahy-a-10-0-UI 5', mRNA sequence.  
 ACCESSION BM722650  
 VERSION 1  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 586)  
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene

JOURNAL  
MEDLINE  
PUBMED  
COMMENT

discovery  
 Genome Res. 6 (9), 791-806 (1996)  
 97044477  
 8889548

Contact: Soares, MB  
 Coordinated Laboratory for Computational Genomics  
 University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565

Email: bento-soares@uiowa.edu

Tissue Procurement: Dr. Gregg Hageman  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research  
 Genetics (www.resgen.com).

Seq primer: M13 Reverse.

## FEATURES

Location/Qualifiers

1..586

/organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-E-E00-ahy-a-10-0-UI"  
 /tissue\_type="fetal eye"  
 /dev\_stage="fetal"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /clone\_lib="UI-E-E00"  
 /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-E00 is a cDNA library containing the following tissue(s): fetal eye. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CGGTATACC. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

## ORIGIN

Alignment Scores:  
 Pred. No.: 1,83e-66 Length: 586  
 Score: 626.00 Matches: 123  
 Percent Similarity: 99.19% Conservativeness: 0  
 Best Local Similarity: 99.19% Mismatches: 0  
 Query Match: 97.97% Indels: 1  
 DB: 4 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x BM722650 (1-586)

QY 1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20  
 DB 200 TCCCTCTTCCGACCGTGTGACCTCCGACATCTCTGAGGTTCGATCGACCGGACAAG 259  
 QY 21 PheValIlePheLeuAspVallyshisPheSerProGluAspLeuThrVallyshisValGln 40  
 DB 260 TTGCTCATCTCTCTGATGTGAGACATCTCTCCCGAGGAGCTCCTGAGTGGAGTGACG 319  
 QY 41 AspAspPheValGluIleHisGlylyshisAsnGluArgGlnAspAspHisGlyTyrIle 60  
 DB 320 GACGACTTTGTGGAGATCCAGGAAGCACACAGGCGCCAGGACGACCGCTTACATT 379  
 QY 61 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 80  
 DB 380 TCCCGTGAGTTCCACCGCGCTTACCGCTCCGCTCAACGTCGACGACGCTCGCCCTCTCT 439  
 QY 81 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLyshisGlnThrGlyLeu 100

```

Db      440 TGCTCCCTGTCCTCCGATGCATGCTGACCTTCTGTGGCCCCAAGATCCAGCTGGCCTG 499
QY      101 AspAlaThr-HisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAl 120
         |||||
Db      500 GATGCCACCCACGCCGAGGAGCCATCCCTGTCGCGGAGGAGAGCCACCTCGGC 559
QY      120 aProSerSer 123
         |||||
Db      560 TCCCTCGTCC 569

RESULT 47
BM696670
LOCUS   607 bp mRNA linear EST 28-FEB-2002
DEFINITION
UI-E-DW0-agk-o-16-0-UI.r1 UI-E-DW0 Homo sapiens cDNA clone
BM696670
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 607)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
8889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 Reverse.

FEATURES
source
1. .607
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-DW0-agk-o-16-0-UI"
/tissue_type="lens"
/dev_stage="adult"
/lab_host="PH108 (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-DW0"
/note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-E-DW0 is a cDNA library containing the following
tissue(s): lens. The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was ligated to an EcoR I adaptor, digested with Not
I, and cloned directionally into pT7T3-Pac vector. The
oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tag for this library is CGATTACGGA. This library
was created for the program, Gene Discovery in the Visual
System, supported by National Eye Institute (NEI)."
```

## ORIGIN

Alignment Scores:  
Pred. No.: 1.92e-66 Length: 607  
Score: 626.00 Matches: 123

```

Percent Similarity: 99.19% Conservative: 0
Best Local Similarity: 99.19% Mismatches: 0
Query Match: 97.97% Indels: 1
DB: 4 Gaps: 0

US-10-657-740-1_COPY_51_173 (1-123) x BM696670 (1-607)

QY      1 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 20
         |||||
Db      84 TCCCTCTTCCGACCCGCTGCTGACTCCGGCATCTCTGAGTTCCGATCCGACCGGACAAG 143
QY      21 PheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrValLysValGln 40
         |||||
Db      144 TTTCGTCACTTCTCGATGTGAAGCACCTTCTCCCGAGGAGACCTCACCTGAAGGTGCG 203
QY      41 AspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTrpIle 60
         |||||
Db      204 GACGACTTTGTGGAGATCCACGAAAGCACAAACAGCGCCAGGACGACCGGTACATT 263
QY      61 SerArgGlu-PheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSe 80
         |||||
Db      264 TCCCGTGAAGTTCCACCGCGCTACCGCTGCCCTCCAACTGACGAGGAGGAGGAGGAGGAG 323
QY      80 rCysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLe 100
         |||||
Db      324 TTGCTCCCTGCTCTCCGATGCGATGCTGACCTTCTGTGGCCCCAAGATCCAGACTGGCCT 383
QY      100 uAspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAl 120
         |||||
Db      384 GGATGCCACCCACGCCGAGGAGCCATCCCTGTCGCGGAGGAGAGAGAGAGAGAGAGAGAG 443
QY      120 aProSerSer 123
         |||||
Db      444 TCCCTCGTCC 453
```

```

RESULT 48
BM706160
LOCUS   512 bp mRNA linear EST 28-FEB-2002
DEFINITION
UI-E-DW0-agg-f-12-0-UI.r1 UI-E-DW0 Homo sapiens cDNA clone
BM706160
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 512)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
8889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 Reverse.

FEATURES
source
1. .512
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
```

/clone="UI-E-DW0-agg-f-12-0-UI"

/tissue\_type="lens"

/dev\_stage="adult"

/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"

/clone\_lib="UI-E-DW0"

/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoR I; Site\_2: Not I; UI-E-DW0 is a cDNA library containing the following tissue(s): lens. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CGATTAGCGA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

#### ORIGIN

Alignment Scores:  
Pred. No.: 8 4e-66 Length: 512  
Score: 620.00 Matches: 119  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 97.03% Indels: 0  
DB: 4 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x BM706160 (1-512)

Qy 5 ThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePhe 24  
Db 3 ACCGTCCTGGACTCCGGCATCTCTGAGGTTCCGATCCGACCGGACAAAGTTCGTCATCTTC 62  
Qy 25 LeuAspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheVal 44  
Db 63 CTCGATGTGAAGCACTTCTCCCGGAGGACCTCACCGTGAAGGTCCAGGACGACTTTGTG 122  
Qy 45 GluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPhe 64  
Db 123 GAGATCCACGGAAAGCACACGAGCGCCAGGACGACCGCTACATTTCCGTCGAGTTC 182  
Qy 65 HisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSer 84  
Db 183 CACCGCCCTACCGCTCCGCTCCAACTGTCGACCACTGCGCCCTCTCTTGTCTCTCTCT 242  
Qy 85 AlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHis 104  
Db 243 GCCGATGCGATGCTGACCTTCTGTGGCCCAAGATCCAGACTGGGCTGGATGCCACCCAC 302  
Qy 105 AlaGluArgAlaIleProValSerArgGluGluLysProThrSerAlaProSerSer 123  
Db 303 GCCGAGCGAGGCCATCCCGTGTGCGGGAGGAGAACCCACCTCGGCTCCCTCGTCC 359

#### RESULT 49

BF726363 544 bp mRNA linear EST 05-JAN-2001  
LOCUS sapiens cDNA clone by05e10 5', mRNA sequence.  
DEFINITION

by05e10.y1 Human Lens cDNA (Un-normalized, unamplified): BY Homo

BF726363

BF726363.1 GI:12042274

EST.

KEYWORDS Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 544)

Wistow G.J., Bernstein S., Behal A. and Smith, D.

NETBANK: EST analysis and bioinformatics for ocular genomics

Invest. Ophthalmol. Vis. Sci. 41 (2000) In press

CONTACT: Wistow G

#### Section on Molecular Structure and Function

National Eye Institute

6/331, NIH, Bethesda, MD 20892-2740, USA

Tel: 301 402 3452

Fax: 301 496 0078

Email: graeme@helix.nih.gov

Plate: 05 row: e column: 10

Seq primer: M13RP1 reverse primer (ABI).

#### FEATURES

source

1..544

Location/Qualifiers

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="by05e10"

/tissue\_type="Lens"

/dev\_stage="Adult"

/lab\_host="EMDH10B"

/clone\_lib="Human Lens cDNA (Un-normalized, unamplified):

By"

/note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses

from different adults (both approximately 40 years old)

together yielded 20ug of total RNA and 150ng mRNA for cDNA

library synthesis. A directionally cloned cDNA library in

the pCMVSPORT6 vector was constructed at Life

Technologies, essentially following the protocols of the

Superscript Plasmid System full details of which are

contained in the manufacturer's instruction manual

(http://www.lifetech.com/). First strand synthesis was

carried out using a Not I primer-adaptor

[5'-pGACTAGTCTAGATCGGACGCGCCCTT15-3']. Not I/blunt

end inserts were cloned into the Not I/EcoR V sites in the

vector. EST analysis was performed on the unamplified

library at the NIH Intramural Sequencing Center (NISC)."

#### ORIGIN

Alignment Scores:  
Pred. No.: 9 09e-66 Length: 544  
Score: 620.00 Matches: 119  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 97.03% Indels: 0  
DB: 2 Gaps: 0

US-10-657-740-1\_COPY\_51\_173 (1-123) x BF726363 (1-544)

Qy 5 ThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePhe 24  
Db 3 ACCGTCCTGGACTCCGGCATCTCTGAGGTTCCGATCCGACCGGACAAAGTTCGTCATCTTC 62  
Qy 25 LeuAspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheVal 44  
Db 63 CTCGATGTGAAGCACTTCTCCCGGAGGACCTCACCGTGAAGGTCCAGGACGACTTTGTG 122  
Qy 45 GluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPhe 64  
Db 123 GAGATCCACGGAAAGCACACGAGCGCCAGGACGACCGCTACATTTCCGTCGAGTTC 182  
Qy 65 HisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSer 84  
Db 183 CACCGCCCTACCGCTCCGCTCCAACTGTCGACCACTGCGCCCTCTCTTGTCTCTCTCT 242  
Qy 85 AlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHis 104  
Db 243 GCCGATGCGATGCTGACCTTCTGTGGCCCAAGATCCAGACTGGGCTGGATGCCACCCAC 302  
Qy 105 AlaGluArgAlaIleProValSerArgGluGluLysProThrSerAlaProSerSer 123  
Db 303 GCCGAGCGAGGCCATCCCGTGTGCGGGAGGAGAACCCACCTCGGCTCCCTCGTCC 359

#### RESULT 50

BM696853

LOCUS

DEFINITION

UI-E-DW0-agg-k-14-0-UI.r1 UI-E-DW0 Homo sapiens cDNA clone

498 bp mRNA linear EST 28-FEB-2002

UI-E-DW0-agl-k-14-0-UI 5', mRNA sequence.

ACCESSION  
BM696853  
VERSION  
BM696853.1 GI:19010111  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM

REFERENCE  
AUTHORS  
TITLE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 498)  
Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene  
discovery

JOURNAL  
MEDLINE  
PUBMED  
COMMENT

Genome Res. 6 (9), 791-806 (1996)  
97044477  
8889548  
Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@uiowa.edu

Tissue Procurement: Dr. Gregg Hageman  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics (www.resgen.com).  
Seq primer: M13 Reverse.

FEATURES  
source

1..498  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-E-DW0-agl-k-14-0-UI"  
/tissue\_type="lens"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone\_lib="UI-E-DW0"  
/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a  
modified polylinker; Site 1: EcoR I; Site 2: Not I;  
UI-E-DW0 is a cDNA library containing the following  
tissue(s): lens. The library was constructed according to  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. First strand cDNA synthesis was primed with an  
oligo-dT primer containing a Not I site. Double stranded  
cDNA was ligated to an EcoR I adaptor, digested with Not  
I, and cloned directionally into pT73-Pac vector. The  
oligonucleotide used to prime the synthesis of  
first-strand cDNA contains a library tag sequence that is  
located between the Not I site and the (dT)18 tail. The  
sequence tag for this library is CGATTAGCGA. This library  
was created for the program, Gene Discovery in the Visual  
System, supported by National Eye Institute (NEI)."

ORIGIN

Alignment Scores:

Pred. No.:	2,51e-65	Length:	498
Score:	616.00	Matches:	118
Percent Similarity:	99.16%	Conservative:	0
Best Local Similarity:	99.16%	Mismatches:	1
Query Match:	96.40%	Indels:	0
DB:	4	Gaps:	0

US-10-657-740-1\_COPY\_51\_173 (1-123) x BM696853 (1-498)

QY	1	SerLeuPheArgThrValLeuAspSerGlyLeuSerGluValArgSerAspAspLys	20
Db	140	TCCTCTTTCCGACCGCTCGGACTTCGGCATCTCTGAGGTTGATCCGACCGGACAG	199
QY	21	PheValIlePheLeuAspValIleHisPheSerProGluAspLeuThrValIleValGln	40
Db	200	TTCTGTCATCTTCTTCGATGTGAGACACTTCTCCCGGAGGAGGACCTCACCGTGAAGGTGAC	259

QY	41	AspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIle	60
Db	260	GACGACTTTGTGGAGATCCACGGAAGCACAAACGAGCGCCAGGACGACCGCTACATT	319
QY	61	SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer	80
Db	320	TCCCGTGAGTTCCACCGCGCTACCGCTCCACGTCGCGTCAACGTGGACCATCGGNCCTCTCT	379
QY	81	CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu	100
Db	380	TGCTCCCTCTGTCGCCGATGGCATGCTGACCTTCTGTGCCCCCAGATCCAGACTGGCCTG	439
QY	101	AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSer	119
Db	440	GATGCCACCCACGCGGAGGAGCCATCCCCCTGTTCGCGGAGGAGAGAGCCACCTCG	496

Search completed: May 30, 2005, 08:48:01  
Job time : 2324.47 secs